Econometric Analysis (700661)

Dr. Keshab Bhattarai*
University of Hull Business School

January 30, 2023

Abstract

Econometric techniques are changing very rapidly along with developments in economic theories. Advanced level of economic analysis requires training in advanced econometrics. Most important theoretical developments in econometrics including fundamental techniques required for specifying, estimating, testing and applying sophisticated time series, cross section or panel data models are presented here for advanced students in econometrics. Solving original problems, reading and reviewing articles to relate findings in refereed journals should complement exercises in this workbook to achieve the learning objectives.

JEL Classification: C

Keywords: Econometrics, Time series, Cross Section, Panel, Bayesian, VAR-Cointegration

*Business School, University of Hull, HU6 7RX, Hull, UK. email: K.R.Bhattarai@hull.ac.uk

Contents

1 L1: Basics of Linear Regression Model

Theories and application of econometrics have evolved over time to analyse econmic relations with cross section,time series and panel data modelling techniques. Classical and Bayesian methods have been applied widely in estimating parameters in single and multiple equation models with the OLS, maximum likelihood, GMM and more recently with non-parametric analysis. While tests for unit roots, Granger causality, cointegration, ARMA, ARIMA and VAR models are essential for time series analysis, cross section models require probit, logit, tobit estimations, duration analysis. Single equation or system methods for static and dynamic panel are becoming increasingly relevant in all areas of economics. BVAR models are becoming more popular in DSGE modelling in recent years. Contributions of prominent economists including Fisher (1923), Durbin and Watson (1950), Klein (1956), McFadden (1963), Box and Jenkins (1976), Heckman (1979), Sims (1980), Hansen (1982), Doan, Litterman and Sims (1984), Phillips (1987), Engle and Granger (1987), Lancaster (1990), Pesaran and Smith (1995), Hendry (1997), Smith (1997), Wooldridge (2002), Chesher (2010) and others made this development possible. Scope of economtrics is expanding very rapidly every year.

- Econometric theory has evolved over time from contributions of many eminent economists including the **econometrics methods and cross section analysis** contained in Fisher (1923), Cochrane and Orcutt (1949), Durbin and Watson (1950), Theil (1956), Klein and Nakamura (1962),McFadden (1963), Farrar and Glauber (1967), Glejser (1969),Ramsey (1969), Rao (1972), Kadane and Anderson (1977), Heckman(1978), Hausman (1978), Blomquist (1980), Hansen (1982) Godfreyand Wickens (1981), Chesher (1984), Zellner (1985), Davidson and Mackinnon (1985),Kiviet (1986),Smith (1987, 1997), White (1987), Wooldridge (1994), Staigler and Stock (1997), Blundell and Preston (1998), Hall (2000), Chesher (2010).

- **time sereis techniques illustrated in** Klein (1956), Box and Jenkins (1976), Hamilton(1994), Harvey (1976), Dickey and Fuller (1979), Hendry(1995), Engle (1982), Engle and Granger (1987), Phillips (1987), Stock and Watson (2002), Nelson and Plosser (1982), Pagan and Wickens (1989), Pyndick and Rubinfeld (1998), Wooldridge (1994) , Enders (2010) Sims (1980), Beveridge and Nelson (1981), Pesaran (1982) Johansen (1988), Baltagi Badi H. (1992), Pesaran and Smith (1995), Garratt, Lee, Phillips (2003) Pesaran and Shin (2003), Hendry (1997), Mills , Pelloni, Zervoyianni (1995), Nelson (1987), Stock and Watson (2001).

- **Bayesian methods developed in** Lancaster(1979), Lancaster and Chesher (1983), Imbens and Lancaster (1994), Bauwens, Lubrano and Richard (1999), Koop (2003), Anscombe (1961),Pratt (1965), Doan, Litterman and Sims (1984), Berger (1990), Chib (1993), Rust (1996) Phillips and Ploberger (1996) Bauwens, Lubrano and Richard (1999) Judge,Griffiths, Hill, Lutkepohl and Lee (1990) Geweke and Keane (2000), Chib, Nardarib and Shephard (2002), Heckelei and Mittelhammer (2003) Canova and Ciccarelli (2004),George, Sun and Ni (2008), Levine, Pearlman, Perendia and Yang (2012).

- **panel date techniques improved** by Wallace and Hussain (1969), Balestra and Nerlove (1966), Hausman (1978), Chamberlain (1984), Arulampalam and Booth(1998), Blundell and Smith (1989),Chesher (1984) , Hansen (1982), Hausman (1978), Heckman (1979), Im, Pesaran and Shin (2003), Imbens and Lancaster (1994), Keifer (1988), Kao (1999), Kwaitkowski, Phillips, Schmidt and Shin (1992), Larsson, Lyhagen and Lothgren (2001) Levin, Lin and Chu (2002), Pedroni (1999), Pesaran and Smith (1995) Phillips (1987), McCoskey and Kao

(1999), Johansen Soren (1988), Johansen Soren (1988) Staigler Stock (1997), Lancaster (1979) Lancaster and Chesher (1983) Zellner A. (1985), Weidmeijer (2005);

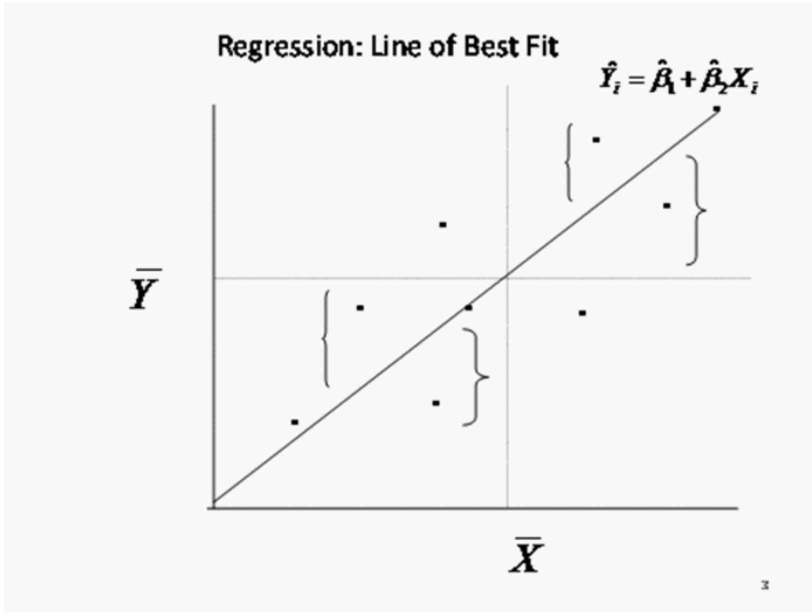
- **macroeconometric models developed by** Hicks (1937), Stone (1942-43), Meade(1951, 1956), Meade et al. (1978), Mirrlees (1971), Pissarides (1984), Mirrlees et al. (2011), Goodhart (1989), Hendry (1997), Blake, Weale (1998), Holly and Weale (2000), King (2004), Bean (1998, 2009), Hendry (1997), Gilhooly, Weale and Wieladek (2012) Wilson (1949), Ash and Smyth (1973), Desai and Weber (1988) and Wallis (1989) were early studies for the UK. Contributions by Mundell (1962), Fleming (1962), Fry and Lilien (1986), Cook, Holly and Turner (2000), Greensdale, Hall, Henry and Nixon (2000), Mellis and Whittaker (2000), Leith and Wren-Lewis (2000), Fisher and Whitley (2000); Blake, Weale and Young (2000) in Holly and Weale (2000); Church, Mitchel, Sault and Wallis (1997), Bean (1998, 2009), Hendry and Clement (2000) Garratt, Lee,Pesaran and Shin (2003), Berentsen, Camera, Waller (2007), Benhabib and Eusepi (2005), Ellison and Pearlman (2011), Driscoll et al. (1983); Den Haan and Marcet (1990),Price (1997), Holland and Scott (1998), Gai, Kapadia, Millard and Perez (2008), Liu and Mumtaz (2011);
- Similarly there are number of excellent texts and eBooks Baltagi (1995), Davidson and MacKinnon(2004) ,Greene (2000), Hsiao (1993) Lancaster (1990), Ruud (2000) Verbeek (2004) Wooldridge (2002) (see detailed list at the end). This section will reviews basic concepts in econometrics. It will review underlying principles of ordinary least square estimators and properties and potential problems and remedial measures to solve above problems. Each concept is supported by examples and dataset that accompany it. Students are expected to look into archives of standard journals to update by replicating results in articles as far as practicable.
- Econometrica (www.econometricsociety.org)
- Econometric Journal (http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1368-423X)
- Journal of Econometrics (http://www.sciencedirect.com/science/journal/03044076)
- Journal of applied econometrics (http://qed.econ.queensu.ca/jae/)
- Phillips P.C.B. (2003) Laws and Limits of Econometrics, Economic Journal, 113, 486, C26-C52
- Hendry, D. F. (1980). 'Econometrics: alchemy or science?', Economica, 47, 387-406. "The three golden rules of econometrics are test, test and test".

A number of software including Excel, OX-GiveWin/PcGive/STAMP, Eviews, Shazam, microfit, JMulti, RATS, LIMDEP, GAUSS, STATA/SPSS, Dynare, GAMS/MPSGE are used to make estimations. More details can be found at:

- **http://www.feweb.vu.nl/econometriclinks/**
- https://www.aeaweb.org/rfe/

Consider a linear regression model:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad i = 1...N \tag{1}$$



- Errors represent all missing elements from this relationship; plus and minuses cancel out resulting in zero mean. Errors are random therefore has constant variance.

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \tag{2}$$

- Normal equations of above regression

$$\sum Y_i = \widehat{\beta}_1 N + \widehat{\beta}_2 \sum X_i \tag{3}$$

$$\sum Y_i X_i = \widehat{\beta}_1 \sum X_i + \widehat{\beta}_2 \sum X_i^2 \tag{4}$$

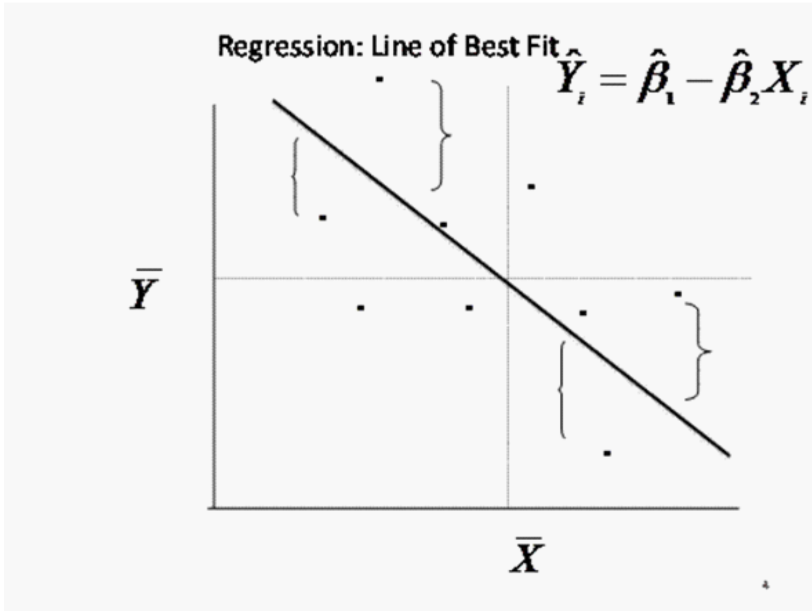
Each dot in the above graph represents an observation. Some observations lie above the least square $\widehat{Y}_i$ line and other observations lie below it. These errors represent all sorts of elements missing from this relationship. Some of them might be due to the missing variables, others might be due to measurement errors, still other may be from the mis-specification of the relationship. The least square line is the line best fits the data set. Differences between each observation and the line $\widehat{Y}_i$ is represented by error terms e_i. As some of them are above the line and others below the line, positive errors cancel out with the negative errors. Note that the least square line passes through the average values of variables X and Y .

1.1 Ordinary Least Square (OLS)

1.1.1 Assumptions

List the OLS assumptions on error terms e_i .

Normality of Errors



$$E(\varepsilon_i) \quad = 0 \tag{5}$$

Homoskedasticity

$$var(\varepsilon_i) \quad = \sigma^2 \qquad for \ \ \forall \ \ i \tag{6}$$

No autocorrelation

$$covar(\varepsilon_i \varepsilon_j) \quad = 0 \tag{7}$$

Independence of errors from dependent variables

$$covar(\varepsilon_i X_i) \quad = 0 \tag{8}$$

1.1.2 Derivation of normal equations for the OLS estimators

Choose $\widehat{\beta}_1$ and $\widehat{\beta}_2$ to minimise sum of square errors:

$$\underset{\widehat{\beta}_1 \widehat{\beta}_2}{Min}\, S = \sum \varepsilon_i^2 = \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{1,i}\right)^2 \tag{9}$$

First order conditions

$$\frac{\partial S}{\partial \widehat{\beta}_1} = 0; \frac{\partial S}{\partial \widehat{\beta}_2} = 0; \tag{10}$$

$$\sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\right)(-1) = 0 \tag{11}$$

$$\sum\left(Y_i-\widehat{\beta}_1-\widehat{\beta}_2X_i\right)(-X_i)=0 \tag{12}$$

$$\sum Y_i=\widehat{\beta}_1N+\widehat{\beta}_2\sum X_i \tag{13}$$

$$\sum Y_iX_i=\widehat{\beta}_1\sum X_i+\widehat{\beta}_2\sum X_i^2 \tag{14}$$

There are two unknown $\widehat{\beta}_1$ and $\widehat{\beta}_2$ and two equations. One way to find $\widehat{\beta}_1$and $\widehat{\beta}_2$ is to use substitution and reduced form method.

Slope estimator by the reduced form equation method; Multiply the second equation by N and first by $\sum X_i$

$$\sum X_i\sum Y_i=\widehat{\beta}_1N\sum X_i+\widehat{\beta}_2\left(\sum X_i\right)^2 \tag{15}$$

$$N\sum Y_iX_i=\widehat{\beta}_1N\sum X_i+\widehat{\beta}_2N\sum X_i^2 \tag{16}$$

By subtraction this reduces to

$$\sum X_i\sum Y_i-N\sum Y_iX_i=\widehat{\beta}_2\left(\sum X_i\right)^2-\widehat{\beta}_2\sum X_i^2 \tag{17}$$

$$\widehat{\beta}_2=\frac{\sum X_i\sum Y_i-N\sum Y_iX_i}{\left(\sum X_i\right)^2-N\sum X_i^2}=\frac{\sum x_iy_i}{\sum x_i^2} \tag{18}$$

This is the OLS Estimator of $\widehat{\beta}_2$, the slope parameter.

Intercept estimator by the reduced form equation method; When $\widehat{\beta}_2$ is known it is easy to find $\widehat{\beta}_1$ by averaging out the regression $Y_i=\beta_1+\beta_2X_i+\varepsilon_i$ as:

$$\widehat{\beta}_1=\overline{Y}-\widehat{\beta}_2\overline{X} \tag{19}$$

Proof:
$\frac{\sum X_i\sum Y_i-N\sum Y_iX_i}{(\sum X_i)^2-N\sum X_i^2}=\frac{\sum x_iy_i}{\sum x_i^2}$;

$$\begin{aligned} LHS &= \frac{\sum X_i\sum Y_i-N\sum Y_iX_i}{\left(\sum X_i\right)^2-N\sum X_i^2}=\frac{N\overline{X}N\overline{Y}-N\sum Y_iX_i}{\left(N\overline{X}\right)^2-N\sum X_i^2} \\ &= \frac{N\overline{X}N\overline{Y}-N\sum Y_iX_i}{\left(N\overline{X}\right)^2-N\sum X_i^2}=\frac{N\overline{XY}-\sum Y_iX_i}{N\overline{X}^2-\sum X_i^2}=\frac{\sum Y_iX_i\ -\ N\overline{XY}}{\sum X_i^2\ -\ N\overline{X}^2} \\ &= \frac{\left(\sum Y_i\text{ - }\overline{Y}\right)\left(\sum X_i\text{ - }\overline{X}\right)}{\left(\sum X_i\text{ - }\overline{X}\right)^2}=\frac{\sum x_iy_i}{\sum x_i^2}=RHS \end{aligned} \tag{20}$$

1.1.3 Normal equations in matrix form

$$Y = XB + e \tag{21}$$

$$\sum Y_i = \widehat{\beta}_1 N + \widehat{\beta}_2 \sum X_i \tag{22}$$

$$\sum Y_i X_i = \widehat{\beta}_1 \sum X_i + \widehat{\beta}_2 \sum X_i^2 \tag{23}$$

$$\begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix}; \quad \widehat{\beta} = (X'X)^{-1} X'Y \tag{24}$$

1.1.4 Estimators:

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix} \tag{25}$$

1.1.5 Data table: an example

DATA		
y	Contant	x
4	1	5
6	1	8
7	1	10
8	1	12
11	1	14
15	1	17
18	1	20
22	1	25

$$\Longrightarrow \begin{bmatrix} \text{Y} \\ 4 \\ 6 \\ 7 \\ 8 \\ 11 \\ 15 \\ 18 \\ 22 \end{bmatrix} = \begin{bmatrix} \multicolumn{2}{c}{\text{X}} \\ 1 & 5 \\ 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \\ 1 & 17 \\ 1 & 20 \\ 1 & 25 \end{bmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} + \begin{bmatrix} e \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

Derivation of OLS Estimators

Matrix multiplication: $\left[(X\prime X) = \begin{matrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{matrix} \right]$

$$(X'X) = \underset{2\times 8}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 8 & 10 & 12 & 14 & 17 & 20 & 25 \end{bmatrix}} \underset{8\times 2}{\begin{bmatrix} 1 & 5 \\ 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \\ 1 & 17 \\ 1 & 20 \\ 1 & 25 \end{bmatrix}} = \underset{2\times 2}{\begin{bmatrix} 8 & 111 \\ 111 & 1843 \end{bmatrix}} \tag{26}$$

OLS in Matrix

$$(X'Y) = \underset{2\times 8}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 5 & 8 & 10 & 12 & 14 & 17 & 20 & 25 \end{bmatrix}} \underset{8\times 1}{\begin{bmatrix} 4 \\ 6 \\ 7 \\ 8 \\ 11 \\ 15 \\ 18 \\ 22 \end{bmatrix}} = \underset{2\times 1}{\begin{bmatrix} 91 \\ 1553 \end{bmatrix}} \tag{27}$$

1.1.6 Summary of data

$$\begin{bmatrix} \sum Y_i = 91 \\ \sum Y_i X_i = 1553 \end{bmatrix} = \begin{bmatrix} N = 8 & \sum X_i = 111 \\ \sum X_i = 111 & \sum X_i^2 = 1843 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \tag{28}$$

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 8 & 111 \\ 111 & 1843 \end{bmatrix}^{-1} \begin{bmatrix} 91 \\ 1553 \end{bmatrix} \tag{29}$$

Solving by the Cramer Rule
Determinant (cross-product)

$$|X'X| = \begin{vmatrix} 8 & 111 \\ 111 & 1843 \end{vmatrix} = (8 \times 1843) - (111 \times 111) = 2423 \tag{30}$$

1.1.7 Estimates

$$\widehat{\beta}_1 = \frac{1}{2423} \begin{vmatrix} 91 & 111 \\ 1553 & 1843 \end{vmatrix} = \frac{167713 - 172383}{2423} = \frac{-4670}{2423} = -1.9274 \tag{31}$$

$$\widehat{\beta}_2 = \frac{1}{2423} \begin{vmatrix} 8 & 91 \\ 111 & 1553 \end{vmatrix} = \frac{12424 - 10101}{2423} = \frac{2323}{2423} = 0.9587 \tag{32}$$

1.1.8 Predicted Y

$$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i \implies \widehat{Y}_i = -1.9274 + 0.9587 X_i \tag{33}$$

Both slope and intercepts make economic sense. In this sample expenditure on foods is determined by weekly income of an individual, people spend 95.6% percent of their weekly income in food expenditure. People who do not have any income receive a income subsidy of 1.93 pence per week.

- Mean prediction

We can use equation this estimate to find the predicted values for each observation on . These are reported as YPRED in the above table. If the weekly income is 40 predicted food expenditure will be 36.42.

$\widehat{Y}_i = -1.9274 + 0.9587X_i = -1.9274 + 0.9587\,(40) = 36.42$

Error terms are also estimated using the fact that

$\widehat{e}_i = Y_i - (-1.9274) - 0.9587X_i = Y_i + 1.9274 - 0.9587X_i$

$$\widehat{Y}_1 = -1.9274 + 0.9587\,(5) = 2.866 \tag{34}$$

$$\widehat{Y}_2 = -1.9274 + 0.9587\,(8) = 5.742 \tag{35}$$

$$\widehat{Y}_3 = -1.9274 + 0.9587\,(10) = 7.660 \tag{36}$$

$$\widehat{Y}_4 = -1.9274 + 0.9587\,(12) = 9.577 \tag{37}$$

$$\widehat{Y}_5 = -1.9274 + 0.9587\,(14) = 11.495 \tag{38}$$

$$\widehat{Y}_6 = -1.9274 + 0.9587\,(17) = 14.371 \tag{39}$$

$$\widehat{Y}_7 = -1.9274 + 0.9587\,(20) = 17.247 \tag{40}$$

$$\widehat{Y}_8 = -1.9274 + 0.9587\,(25) = 22.041 \tag{41}$$

1.1.9 Estimated errors

$$\widehat{e}_i = Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \;\; = Y_i - (-1.9274 + 0.9587X_i) \tag{42}$$

$$\widehat{e}_1 = 4 + 1.9274 - 0.9587\,(5) = 1.134 \tag{43}$$

$$\widehat{e}_2 = 6 + 1.9274 - 0.9587\,(8) = 0.258 \tag{44}$$

$$\widehat{e}_3 = 7 + 1.9274 - 0.9587\,(10) = -0.660 \tag{45}$$

$$\widehat{e}_4 = 8 + 1.9274 - 0.9587\,(12) = -1.580 \tag{46}$$

$$\widehat{e}_5 = 11 + 1.9274 - 0.9587\,(14) = -0.495 \tag{47}$$

$$\widehat{e}_6 = 15 + 1.9274 - 0.9587\,(17) = 0.629 \tag{48}$$

$$\widehat{e}_7 = 18 + 1.9274 - 0.9587\,(20) = 0.753 \tag{49}$$

$$\widehat{e}_8 = 22 + 1.9274 - 0.9587\,(25) = 0.000 \tag{50}$$

- Use of regression estimates to calculate the elasticities

The definition of elasticity of food expenditure on income is given by

$$\eta = \frac{\frac{\partial Y}{Y}}{\frac{\partial X}{X}} = 0.0.9587 \times \frac{13.857}{11.375} = 1.1683 \tag{51}$$

This suggests that the expenditure on food is elastic around the mean. There will be 17 pence more expenditure to every £1 rise in weekly income.

$$\begin{aligned} var\,(\,Y_i) &= \sum\left[Y_i - \overline{Y}_i\right]^2 = \sum\left[\widehat{Y}_i - \overline{Y}_i + \,\widehat{e}_i\right]^2 \\ &= \sum\left(\widehat{Y}_i - \overline{Y}_i\right)^2 + \sum\widehat{e}_i^2 + 2\sum\left(\widehat{Y}_i - \overline{Y}_i\right)\widehat{e}_i \\ &= \sum\left(\widehat{Y}_i - \overline{Y}_i\right)^2 + \sum\widehat{e}_i^2 \end{aligned} \tag{52}$$

$$TSS = RSS + ESS \tag{53}$$

For N observations and K explanatory variables
[Total variation] = [Explained variation] + [Residual variation]
df = N-1 K-1 T-K

1.1.10 Variances

$$\begin{aligned} \sum\widehat{e}_i^2 &= \widehat{e}_1^2 + \widehat{e}_2^2 + \widehat{e}_3^2 + \widehat{e}_4^2 + \widehat{e}_5^2 + \widehat{e}_6^2 + \widehat{e}_7^2 + \widehat{e}_8^2 \\ &= \left[\begin{array}{c} (1.134)^2 + (0.258)^2 + (-0.660)^2 + (-1.580)^2 + (-0.495)^2 \\ + (0.629)^2 + (0.753)^2 + (0.000)^2 \end{array}\right] = 5.484 \end{aligned}$$

$$var\,(\,\widehat{e}_i) = E\left(\widehat{\varepsilon}_i^2\right) = \frac{\sum\widehat{e}_i^2}{N-k} = \widehat{\sigma}^2 \tag{54}$$

Where k = number of parameters in the regression; N = number of observations

$$\frac{\sum\widehat{e}_i^2}{N-k} = \frac{5.4841}{8-2} = 0.914 \tag{55}$$

$$\sum y_i^2 = \sum\left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - N\overline{Y}^2 = 1319 - 8\times 11.375^2 = 283.875 \tag{56}$$

$$\sum x_i^2 = \sum\left(X_i - \overline{X}\right)^2 = \sum X_i^2 - N\overline{X}^2 = 1843 - 8\times 13.875^2 = 302.875 \tag{57}$$

1.1.11 R-square and F statistics

$$\sum\widehat{y}_i^2 = \widehat{\beta}_2^2\sum x_i^2 = 0.9587^2\times 302.875 = 278.390 \tag{58}$$

Coefficient of determination (line of best fit)

Coefficient of determination is a measure in the regression analysis that shows the explanatory power of independent variables (regressors) in explaining the variation on dependent variable (regressand). The total variation on the dependent variable can be decomposed as following:

$$R^2 = \frac{\sum\widehat{y}_i^2}{\sum y_i^2} = \frac{278.390}{283.875} = 0.981 \tag{59}$$

For N observations and K explanatory variables
[Total variation] = [Explained variation] + [Residual variation]

df = N-1 K-1 T-K

$$F = \frac{RSS/(K-1)}{ESS/(N-k)} = \frac{\frac{278.390}{1}}{\frac{5.4841}{6}} = \frac{278.390}{0.9140} = 304.579 \tag{60}$$

$$\overline{R}^2 = 1-(1-R^2)\frac{N-1}{N-K} = 1-(1-0.981)\frac{8-1}{8-2} = 0.978 \tag{61}$$

$R^2 > \overline{R}^2$
Prove that two forms $\overline{R}^2 = 1-(1-R^2)\frac{N-1}{N-K}$ or $\overline{R}^2 = R^2\frac{N-1}{N-K} - \frac{K-1}{N-K}$ are equivalent.

Relation between Rsquare and Rbarsquare Prove that two forms $\overline{R}^2 = 1-(1-R^2)\frac{N-1}{N-K}$ or $\overline{R}^2 = R^2\frac{N-1}{N-K} - \frac{K-1}{N-K}$ are equivalent

Proof

$$\begin{aligned}
LHS &= \overline{R}^2 = 1-(1-R^2)\frac{N-1}{N-K} = R^2 + \left(1-R^2\right) - \left(1-R^2\right)\frac{N-1}{N-K} \\
&= R^2 - \left(1-R^2\right)\left[\frac{N-1}{N-K} - 1\right] = R^2 + \left(1-R^2\right)\left[\frac{N-1-N+K}{N-K}\right] \\
&= R^2 - \left(1-R^2\right)\left[\frac{K-1}{N-K}\right] = R^2 + R^2\frac{K-1}{N-K} - \frac{K-1}{N-K} \\
&= R^2\left(1+\frac{K-1}{N-K}\right) - \frac{K-1}{N-K} \\
&= R^2\left(\frac{N-K+K-1}{N-K}\right) - \frac{K-1}{N-K} = R^2\left(\frac{N-1}{N-K}\right) - \frac{K-1}{N-K} \\
& RHS; \quad QED
\end{aligned} \tag{62}$$

1.1.12 Variance, standard error and t-value of slope parameter

$$Var\left(\widehat{\beta}_2\right) = var\left[\frac{\sum\left(X_i - \overline{X}\right)}{\sum\left(X_i - \overline{X}\right)^2}\right] var\left(y_i\right) = \frac{1}{\sum x_i^2}\widehat{\sigma}^2 \tag{63}$$

$$var\left(\widehat{\beta}_2\right) = \frac{0.914}{302.875} = 0.0030 \tag{64}$$

$$SE\left(\widehat{\beta}_2\right) = \sqrt{0.0030} = 0.0548 \tag{65}$$

$$t_{\widehat{\beta}_2} = \frac{\widehat{\beta}_2 - \beta_2}{SE\left(\widehat{\beta}_2\right)} = \frac{0.9587-0}{0.0548} = 17.495 \tag{66}$$

Variance, Standard Error and T value of Intercept Parameter

$$var\left(\widehat{\beta}_1\right) = \left[\frac{1}{N}+\frac{\overline{X}^2}{\sum x_i^2}\right]\widehat{\sigma}^2 \tag{67}$$

$$var\left(\widehat{\beta}_1\right) = \left[\frac{1}{8}+\frac{13.875^2}{302.875}\right]\times 0.914 \tag{68}$$

$$var\left(\widehat{\beta}_1\right) = [0.125+0.634]\times 0.914 = 0.6937 \tag{69}$$

$$SE\left(\widehat{\beta}_1\right) = \sqrt{0.6937} = 0.833 \tag{70}$$

$$t_{\widehat{\beta}_1} = \frac{\widehat{\beta}_1-\beta_1}{SE\left(\widehat{\beta}_1\right)} = \frac{-1.9774-0}{0.833} = -2.374 \tag{71}$$

For a review of matrix algebra see the appendix.

1.1.13 Exercise 1

Regress demand for a product (Y_i) on its own prices (X_i) as following

$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ldots N$$

where e_i is a randomly distributed error term for observation i.

1. (a) List the OLS assumptions on error terms e_i .
   (b) Derive the normal equations and the OLS estimators of $\widehat{\beta}_1$ and $\widehat{\beta}_2$.
   (c) A shopkeeper observed the data quantities and prices as given in Table 2 below. What are the OLS estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ implied by these data? Is this a normal good?
   (d) What are the variances of e_i and Y_i?
   (e) What are R^2 and $\overline{R}^2$?
   (f) Determine the overall significance of this model by F-test at 5 percent level of significance. [Critical value of F for df(1,4) =7.71]
   (g) What are the variances and standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$?
   (h) Compute t-statistics and determine whether parameters $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are statistically significant at 5 percent level of significance.
      i. [Critical value of t for five percent significance for 4 degrees of freedom is 2.776 (i.e $t_{crit,0.05,4} = 2.777$)]
   (i) What is the prediction of Y when X is 0.5?
   (j) What is the elasticity of demand evaluated at the mean values of Y_i and X_i?

Table 1: Data on Quantities and Prices

Quantities (Y_i)	5	10	15	20	25	30
Prices (X_i)	10	8	6	4	2	1

(k) Reformulate the model to include price of a substitute product in the model. What will happen to this estimation if these two prices are exactly correlated?

(l) How would you decide whether demand for this product varies by gender?

Hints: $\left[\sum X_i = 31 \quad \sum X_i^2 = 221 \quad \sum Y_i^2 = 2275; \sum Y_i = 105 \quad \sum Y_i X_i = 380\right]$;
$(X'X)^{-1} = \begin{bmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{bmatrix}$

Test whether work-hours depend on weekly or annual pay among UK counties using data Unempl_pay-couties.csv.

1.2 Statistical inference

What is the statistical inference?

- Inference is statement about population based on sample information.
- Economic theory provides these relations. Statistical inference is about empirically testing whether those relations are true based on available cross section, time series or panel data.
- Hypotheses are set up according to the economic theory, estimates of parameters are estimated using OLS (similar other) estimators.
- Consider a linear regression

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \qquad i = 1 \ldots N \tag{72}$$

Here the true values of β_1 and β_2 are unknown parameters. Their values can be estimated using the OLS technique. $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are such estimates. Validity of these estimates are tested using statistical distributions. Two most important tests for a linear regression are

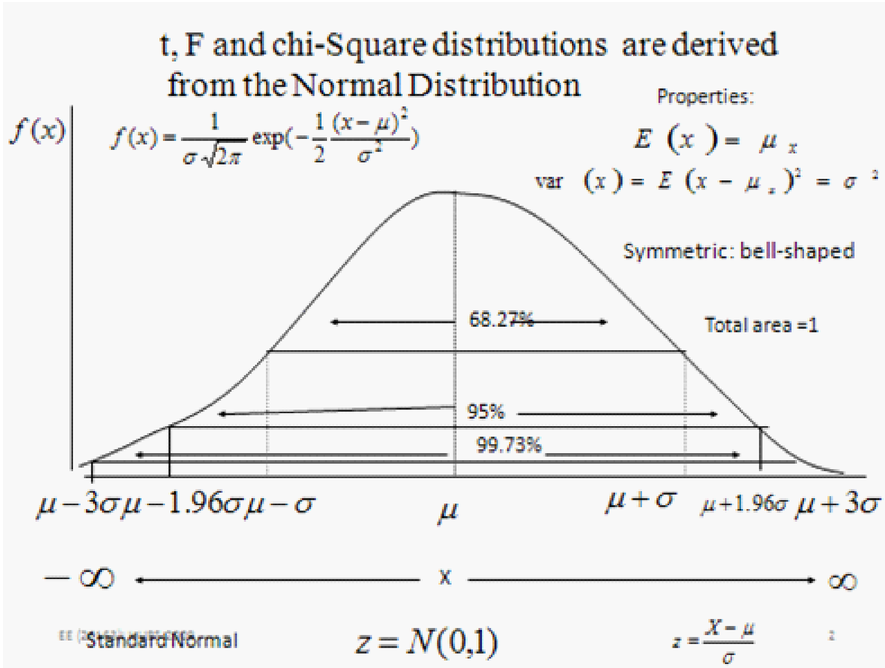
1. Significance of an individual coefficient: **t-test**
2. Overall significance of the model: **F -test**
3. **Overall fit of the data to the model is indicated by** R^2. (χ^2, Durbin-Watson, Unit root tests to be discussed later).

The ordinate at z is given by the standard normal density function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

The probabilities and areas are given by the standard normal distribution function

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt.$$



1.2.1 Hypothesis

Standard hypothesis about individual coefficients (t-test)

Null hypothesis: value of intercept and slope coefficients are zero.

$$H_0 : \beta_1 = 0$$
$$H_0 : \beta_2 = 0$$

Alternative hypotheses: Intercept and slope coefficients are non -zero.

$$H_A : \beta_1 \neq 0$$
$$H_A : \beta_2 \neq 0$$

Parameter β_2 is slope, $\frac{\partial Y}{\partial X}$; it measures how much Y will change when X changes by one unit. Parameter β_1 is intercept. It shows amount of Y when X is zero.

Economic theory: a normal demand function should have $\beta_1 > 0$ and $\beta_2 < 0$; a normal supply function should have $\beta_1 \neq 0$ $\beta_2 > 0$. This is the hypothesis to be tested empirically.

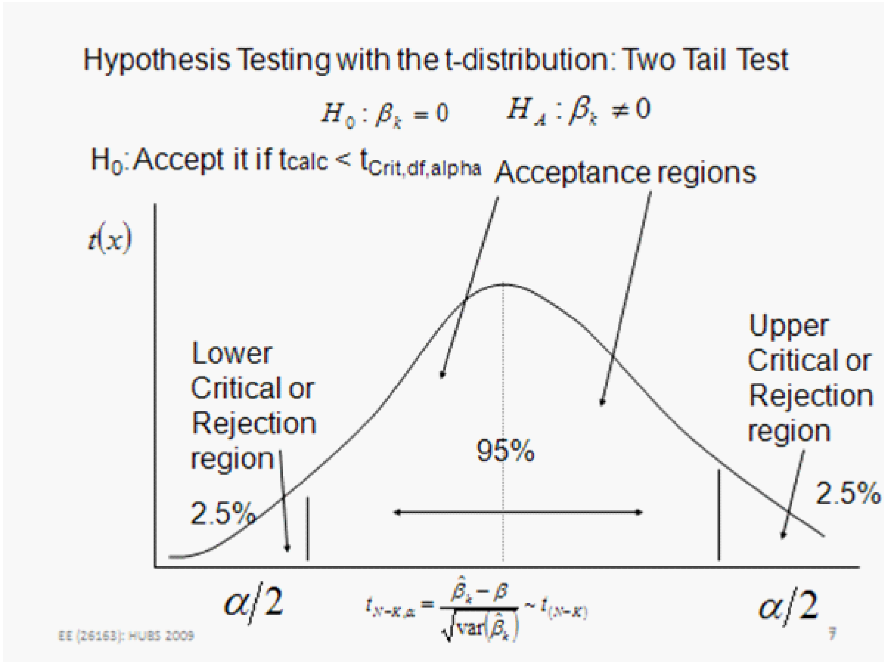
Standard hypothesis about the validity of the model (F-test)

Null hypothesis: both intercept and slope coefficients are zero; model is meaningless and irrelevant:

$$H_0 : \beta_1 = \beta_2 = 0$$

Alternative hypotheses: at least one of the parameters is non -zero, model is relevant:

$$H_A \text{ :either } \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ or both } \beta_1 \neq 0, \beta_2 \neq 0$$



As is often seen, some of the coefficients in a regression model may be insignificant but F-statistics is significant and model is valid.

An Example of regression on deviations from the mean

Table 2: Data Table:Price and Quantity

X	1	2	3	4	5	6
Y	6	3	4	3	2	1

What are the estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$?
Here $\sum X_i = 21$; $\sum Y_i = 19$; $\sum Y_i X_i$ =52 $\sum X_i^2$ =91 $\sum Y_i^2$ =75 $\overline{Y} = 3.17$ $\overline{X} = 3.5$
OLS estimators

$$\widehat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2}; \qquad \widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \tag{73}$$

1.2.2 Normal equations and its deviation form

- Normal equations of above regression

$$\sum Y_i = \widehat{\beta}_1 N + \widehat{\beta}_2 \sum X_i \tag{74}$$

$$\sum Y_i X_i = \widehat{\beta}_1 \sum X_i + \widehat{\beta}_2 \sum X_i^2 \tag{75}$$

Define deviations as

$$x_i = \left(X_i - \overline{X}\right) \tag{76}$$

$$y_i = \left(Y_i - \overline{y}\right) \tag{77}$$

$$\sum \left(X_i - \overline{X}\right) = 0; \sum \left(Y_i - \overline{y}\right) = 0 \tag{78}$$

Normal Equations and Deviation Form
Putting these in the Normal equations

$$\sum \left(Y_i - \overline{y}\right) = \widehat{\beta}_1 N + \widehat{\beta}_2 \sum \left(X_i - \overline{X}\right) \tag{79}$$

$$\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{y}\right) = \widehat{\beta}_1 \sum \left(X_i - \overline{X}\right) + \widehat{\beta}_2 \sum \left(X_i - \overline{X}\right)^2 \tag{80}$$

Terms $\sum \left(X_i - \overline{X}\right) = 0; \sum \left(Y_i - \overline{y}\right) = 0$ drop out $\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{y}\right) = \sum x_i y_i$
and $\sum \left(X_i - \overline{X}\right)^2 = \sum x_i^2$

This is a regression through origin. Therefore estimator of slope ceofficient with deviation

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{81}$$

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \tag{82}$$

- The reliability of $\widehat{\beta}_2$ and $\widehat{\beta}_1$ depends on their variances; t-test is used to determine their significance.

1.2.3 Deviations from the mean

Useful short-cuts (though matrix method is more accurate, sometimes quick short cuts like this can be handy)

$$\sum x_i^2 = \sum \left(X_i - \overline{X}\right)^2 = \sum X_i^2 - N\overline{X}^2 = 91 - 6(3.5)^2 = 17.5 \tag{83}$$

$$\sum y_i^2 = \sum \left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - N\overline{Y}^2 = 91 - 6(3.17)^2 = 14.7 \tag{84}$$

$$\begin{aligned}
\sum y_i x_i &= \sum \left(Y_i - \overline{Y}\right) \sum \left(X_i - \overline{X}\right) \\
&= \sum Y_i X_i - \overline{Y} \sum X_i - \overline{X} \sum Y_i + N\overline{Y}\overline{X} = \\
&\quad \sum Y_i X_i - \overline{Y} N \overline{X} - \overline{X} N \overline{Y} + N \overline{Y}\overline{X} \\
&= \sum Y_i X_i - \overline{Y} N \overline{X} = 52 - (3.5)(6)(3.17) = -14.57
\end{aligned} \tag{85}$$

1.2.4 OLS estimates by the deviation method

Estimate of the slope coefficient:

$$\widehat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{-14.57}{17.5} = -0.833 \tag{86}$$

This is negative as expected.
Estimate of the intercept coefficient.

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} = 3.17 - (-0.833)(3.5) = 6.09 \tag{87}$$

It is positive as expected.
Thus the **regression line fitted from the data**

$$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i = 6.09 - 0.833 X_i \tag{88}$$

How reliable is this line? Answer to this should be based on the analysis of variance and statistical tests.

1.2.5 Variation of Y, predicted Y and error

Total variation to be explained:

$$\sum y_i^2 = \sum \left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - N\overline{Y}^2 = 75 - 6(3.17)^2 = 14.707 \tag{89}$$

Variation explained by regression:

$$\begin{aligned}\sum \widehat{y}_i^2 &= \sum (\widehat{\beta}_2 x_i)^2 = \widehat{\beta}_2^2 \sum {x_i}^2 = \left(\frac{\sum y_i x_i}{\sum x_i^2}\right)^2 \sum {x_i}^2 \\ &= \frac{\left(\sum y_i x_i\right)^2}{\sum x_i^2} = \frac{(-14.57)^2}{17.5} = \frac{212.28}{17.5} = 12.143\end{aligned} \tag{90}$$

Note that in deviation form: $\sum \widehat{y}_i = \sum \widehat{\beta}_2 x_i$.
Unexplained variation (accounted by various errors):

$$\sum \widehat{e}_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2 = 14.707 - 12.143 = 2.564 \tag{91}$$

1.2.6 Measure of Fit: R-square and Rbar-square

The **measure of fit** R^2 is ratio of total variation explained by regression $\left(\sum \widehat{y}_i^2\right)$ to total variation that need to be explained $\left(\sum y_i^2\right)$

$$R^2 = \frac{\sum \widehat{y}_i^2}{\sum y_i^2} = \frac{12.143}{14.707} = 0.826 \tag{92}$$

This regression model **explains about 83 percent** of variation in y.

$$\overline{R}^2 = 1 - (1 - R^2)\frac{N-1}{N-K} = 1 - (1 - 0.826)\frac{5}{4} = 0.78 \tag{93}$$

Variance of error indicates the unexplained variation

$$var\left(\widehat{e}_i\ \right) = \widehat{\sigma}^2 = \frac{\sum \widehat{e}_i^2}{N-K} = \frac{2.564}{4} = 0.641 \tag{94}$$

$$var\left(y_i\ \right) = \frac{\sum y_i^2}{N-1} = \frac{14.7}{5} = 2.94 \tag{95}$$

1.2.7 Variance of parameters

Reliability of estimated parameters depends on their variances, standard errors and t-values

$$var\left(\widehat{\beta}_2\right) = \frac{1}{\sum {x_i}^2}\widehat{\sigma}^2 = \frac{0.641}{17.5} = 0.037 \tag{96}$$

$$var\left(\widehat{\beta}_1\right) = \left[\frac{1}{N} + \frac{\overline{X}^2}{\sum {x_i}^2}\right]\widehat{\sigma}^2 = \left[\frac{1}{6} + \frac{3.5^2}{17.5}\right]0.641 = (0.867)\,0.641 = 0.556 \tag{97}$$

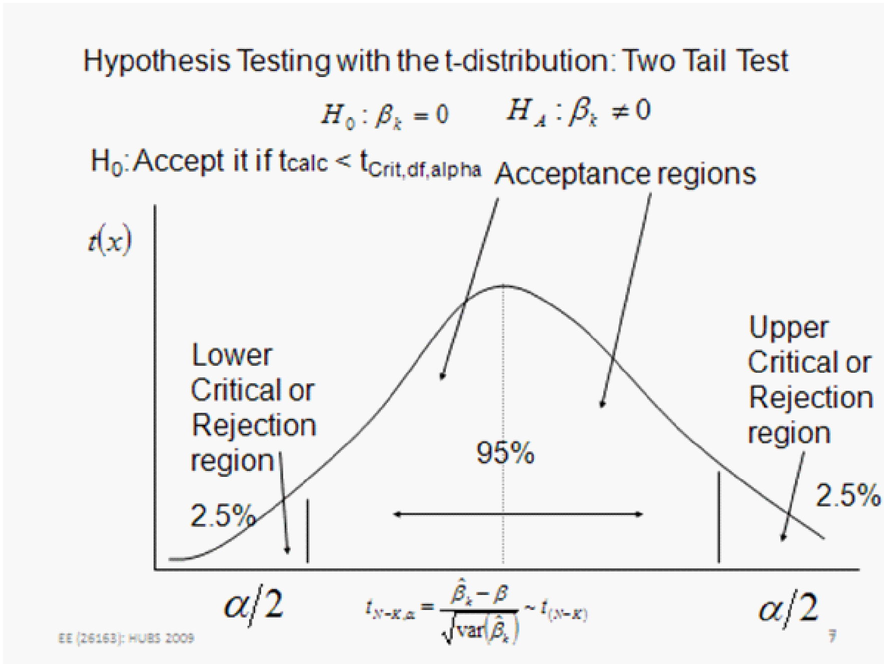
Prove these formula (see later on).

Standard errors

$$SE\left(\widehat{\beta}_2\right) = \sqrt{var\left(\widehat{\beta}_2\right)} = \sqrt{0.037} = 0.192 \tag{98}$$

$$SE\left(\widehat{\beta}_1\right) = \sqrt{var\left(\widehat{\beta}_1\right)} = \sqrt{0.556} = 0.746 \tag{99}$$

1.2.8 t-test



Theoretical value of T distribution is derived by dividing mean by standard error. Mean is a normally distributed variable and the standard error χ^2 distribution. Originally t-distribution was established by W.S. Gossett of Guiness Brewery in 1919.

One- and Two-Tailed Tests

If the area in only one tail of a curve is used in testing a statistical hypothesis, the test is called a **one-tailed test**; if the area of both tails are used, the test is called **two-tailed**.

The decision as to whether a one-tailed or a two-tailed test is to be used depends on the alternative hypothesis.

1.2.9 Test of significance of parameters (t-test)

$$t\left(\widehat{\beta}_2\right) = \frac{\widehat{\beta}_2}{SE\left(\widehat{\beta}_2\right)} = \frac{-0.833}{0.192} = -4.339 \tag{100}$$

$$t\left(\widehat{\beta}_1\right) = \frac{\widehat{\beta}_1}{SE\left(\widehat{\beta}_1\right)} = \frac{6.09}{0.746} = 8.16 \tag{101}$$

These calculated t-values need to be compared to t-values from the theoretical t-table.

Decision rule: (one tail test following economic theory)

- Accept $H_0 : \beta_1 > 0$ if $t\left(\widehat{\beta}_1\right) < t_{\alpha,df}$;
- Reject $H_0 : \beta_1 > 0$ or accept $H_A : \beta_1 \ngtr 0$ if $t\left(\widehat{\beta}_1\right) > t_{\alpha,df}$

- Accept $H_0 : \beta_2 < 0$ if $t\left(\widehat{\beta}_2\right) < t_{\alpha,df}$
- Reject $H_0 : \beta_2 < 0$ or accept $H_A : \beta_2 \nleq 0$ if $t\left(\widehat{\beta}_2\right) > t_{\alpha,df}$

P-value: Probability of test statistics exceeding that of the sample statistics.

Test of significance of parameters (t-test)

Theoretical values of t are given in a t Table. Column of t-table have level of significance (α) and rows have degrees of freedom.

Here $t_{\alpha,df}$ is t-table value for degrees of freedom ($df = n - k$) and α level of significance. df = 6-2=4.

Table 3: Relevant t-values (one tail) fron t-Table

(n,α)	0.05	0.025	0.005
1	6.314	12.706	63.657
2	2.920	4.303	9.925
4	2.132	2.776	4.604

$t\left(\widehat{\beta}_1\right) = 8.16 > t_{\alpha,df} = t_{0.05,4} = 2.132$. Thus the intercept is statistically significant; $t\left(\widehat{\beta}_2\right) = |-4.339| > t_{\alpha,df} = t_{0.05,4} = 2.132$. Thus the slope is also statistically significant at 5% and 2.5% level of significance.

1.2.10 Confidence interval on the slope parameter

A researcher may be interested more in knowing the interval in which the true parameter may lie than in the point estimte where α is the level of significance or the probability of error such as 1% or 5%. That means accuracy of the estimate is $(1-\alpha)$ %.

A 95% level confidence interval for β_1 and β_2 is:

$$P\left[\widehat{\beta}_2 - SE\left(\widehat{\beta}_2\right)t_{\alpha,n} < \beta_2 < \widehat{\beta}_2 + SE\left(\widehat{\beta}_2\right)t_{\alpha,n}\right] = (1-\alpha) \tag{102}$$

$$\begin{aligned} & P\left[-0.833 - 0.192\,(2.132.) < \beta_2 < -0.833 + 0.192\,(2.132.)\right] \\ = \;& (1-0.05) = 0.95 \end{aligned} \tag{103}$$

$$P\left[-1.242 < \beta_2 < -0.424\right] = 0.95 \tag{104}$$

There is 95 confidence that the true value of slope β_2 lies between -0.424 and -1.242.

Confidence interval on the intercept parameter

95 % confidence interval on the slope parameter:

$$P\left[\widehat{\beta}_1 - SE\left(\widehat{\beta}_2\right)t_{\alpha,n} < \beta_1 < \widehat{\beta}_1 + SE\left(\widehat{\beta}_2\right)t_{\alpha,n}\right] = (1-\alpha) \tag{105}$$

$$\begin{aligned} & P\left[6.09 - 0.746\,(2.132.) < \beta_1 < 6.09 + 0.746\,(2.132.)\right] \\ = \;& (1-0.05) = 0.95 \end{aligned} \tag{106}$$

$$P\left[4.500 < \beta_2 < 7.680\right] = 0.95 \tag{107}$$

There is 95 confidence that the true value of intercept β_1 lies between 4.500 and 7.680.

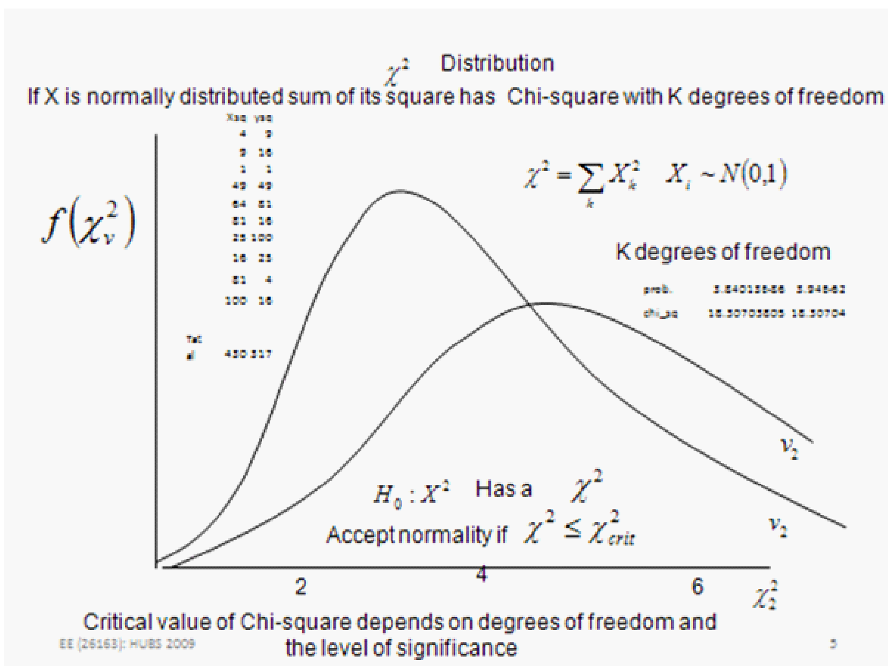
1.2.11 F-test

F-value is the ratio of sum of squared normally distributed variables (χ^2) adjusted for relevant degrees of freedom.

$$F = \frac{V_1/n_1}{V_2/n_2} = F\left(n_1, n_2\right) \tag{108}$$

Where V_1 and V_2 are variances of numberator and denomenator and $n_1$and n_2 are degrees of freedom of numberator and denomenator.

H_0: Variance are the same; H_A: Variance are different. F_{crit} values are obtained from F-distribution table. Accept it if $F_{Calc} < F_{crit}$ and reject if $F_{Calc} > F_{crit}$



F- is ratio of two χ^2 distributed variables with degrees of freedom $n2$ and $n1$.

$$F_{calc} = \frac{\frac{\sum \widehat{y}_i^2}{K-1}}{\frac{\sum \widehat{e}_i^2}{N-K}} = \frac{\frac{12.143}{1}}{\frac{2.564}{4}} = \frac{12.143}{0.641} = 18.94 \tag{109}$$

$n1$ = degrees of freedom of numerator; $n2$ =degrees of freedom of denominator; for 5% level of significance $F_{n1,n2} = F_{1,4} = 7.71$; $F_{calc} > F_{1,4}$;for 1% level of significance $F_{n1,n2} = F_{1,4} = 21.20$;

Table 4: Relevant F-values from the F-Table

	1% level of significance			5% level of significance		
$(n2, n1)$	1	2	3	1	2	3
1	4042	4999.5	5403	161.4	199.5	215.7
2	98.50	99.00	99.17	18.51	19.00	19.16
4	21.20	18.00	16.69	7.71	6.94	6.59

$F_{calc} > F_{1,4} \Longrightarrow$imply that this model is statistically significant at 1% as well as at 5% level of significance. Model is meaningful.

Exercise: Revise data as following and do all above calculations.

Table 5: Data Table:Price and Quantity

X	1	2	3	4	5	6
Y	6	5	4	3	2	1

This should give a line of perfect fit. What does it impy to $\sum \widehat{e}_i^2$?

1.3 Type I and Type II errors

Table 6: Type I and Type II Errors

	True	False
Accept	Correct Decision	Type II Error, (β)
Reject	Type I Error, (α)	Correct Decision

α : level of significance (probability of type I error) β : probability of type II error

- Type II occurs when the Null hypothesis is wrong.
- Power of test: probability of rejecting the null while it is false.
- Power= 1-beta = 1- Prob(type II error)

1.3.1 Prediction and error of prediction

What is the prediction of Y when X is 0.5?

$$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i = 6.09 - 0.833\,(0.5) = 5.673 \tag{110}$$

Prediction error

$$f = Y_0 - \widehat{Y}_0 = \beta_1 + \beta_2 X_i + \varepsilon_0 \;\; - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \tag{111}$$

Mean of prediction error

$$E(f) = E\left(\beta_1 + \beta_2 X_i + \varepsilon_0 \ - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \ \right) = 0 \tag{112}$$

Predictor is ubiased.

1.3.2 t-test for variance of forecast

$$t_f = \frac{Y_0 - \widehat{Y}_0}{SE(f)} \sim t_{N-2} \tag{113}$$

Standard error of forecast. Find $var(f)$.

$$SE(f) = \sqrt{var(f)} \tag{114}$$

Confidence interval of forecast

$$\Pr\left[-t_c \leq \frac{Y_0 - \widehat{Y}_0}{SE(f)} \leq t_c\right] = (1-\alpha) \tag{115}$$

$$\Pr\left[\widehat{Y}_0 - t_c SE(f) \leq Y_0 \leq \widehat{Y}_0 + t_c SE(f)\right] = (1-\alpha) \tag{116}$$

Variance of Y and error

$$E(\widehat{\varepsilon}_i)^2 = \frac{\sum \widehat{e}_i^2}{N-k} = \widehat{\sigma}^2 \tag{117}$$

where N is is number of observations and k is the number of parameters including intercept.

$$\begin{aligned} var(Y_i) &= E\left(Y_i - \overline{Y}\right)^2 = E\left[\beta_1 + \beta_2 X_i + \varepsilon_i - \widehat{\beta}_1 - \widehat{\beta}_2 \overline{X} \ \right]^2 \\ &= \left[\beta_1 + \beta_2 E(X_i) + E(\varepsilon_i) - E\left(\widehat{\beta}_1\right) - E\left(\widehat{\beta}_2\right)\overline{X} \ \right]^2 \\ &= \left[\beta_1 + \beta_2 \overline{X} + E(\varepsilon_i) - \beta_1 - \beta_2 \overline{X} \ \right]^2 \\ &= \left[\beta_1 + \beta_2 \overline{X} + E(\varepsilon_i) - \beta_1 - \beta_2 \overline{X} \ \right]^2 \\ &= \left[E(\varepsilon_i) \ \right]^2 = \sigma^2 \end{aligned} \tag{118}$$

1.3.3 Variance of slope parameter

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{119}$$

$$E\left(\widehat{\beta}_2\right) = \sum w_i y_i \tag{120}$$

where

$$w_i = \frac{x_i}{\sum x_i^2} = \frac{\left(X_i - \overline{X}\right)}{\sum \left(X_i - \overline{X}\right)^2} \tag{121}$$

$$Var\left(\widehat{\beta}_2\right) = var\left[\frac{\sum \left(X_i - \overline{X}\right)}{\sum \left(X_i - \overline{X}\right)^2}\right] var(y_i) = \frac{1}{\sum x_i^2}\widehat{\sigma}^2 \tag{122}$$

1.3.4 Variance of intercept parameter

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \tag{123}$$

$$\begin{aligned}
var\left(\widehat{\beta}_1\right) &= var\left(\overline{Y} - \widehat{\beta}_2 \overline{X}\right) \\
&= E\left[\frac{\sum y_i}{N} - \overline{X}\frac{\sum x_i y_i}{\sum x_i^2}\right]^2 \\
&= E\sum\left[\frac{1}{N} - \overline{X}\frac{x_i}{\sum x_i^2}\right]^2 E\left[\sum y_i\right]^2 \\
&= \left[\frac{N}{N^2} + \overline{X}^2 \sum w_i^2 - 2\frac{1}{N}\overline{X}\sum w_i\right]\widehat{\sigma}^2 &(124) \\
&= \left[\frac{1}{N} + \frac{\overline{X}^2}{\sum x_i^2}\right]\widehat{\sigma}^2 &(125)
\end{aligned}$$

$-2\frac{1}{N}\overline{X}\sum w_i = 0$ because $\sum w_i = 0'$

1.3.5 Covariance of parameters (with matrix)

$$\begin{aligned}
b &= \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(X'X\right)^{-1} X'Y = \left(X'X\right)^{-1} X'\left(X\beta + e\right) \\
&= \beta + \left(X'X\right)^{-1} X'e &(126)
\end{aligned}$$

$$b - \beta = \left(X'X\right)^{-1} X'e \tag{127}$$

$$cov\left(b-\beta\right) = E\left[\left(X'X\right)^{-1} X'ee'X\left(X'X\right)^{-1}\right] = \left(X'X\right)^{-1}\sigma^2 \tag{128}$$

$$\left(X'X\right)^{-1} = \begin{pmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}^{-1} \tag{129}$$

$$cov\left(b-\beta\right) = \left(X'X\right)^{-1}\sigma^2 = \frac{1}{N\sum X_i^2 - \left(\sum X_i\right)^2}\begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & N \end{bmatrix} \tag{130}$$

1.3.6 Covariance of parameters (with matrix)

$$\left(X'X\right)^{-1} = \begin{pmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}^{-1} \tag{131}$$

$$cov\left(b-\beta\right) = \begin{bmatrix} var\left(b_1\right) & var\left(b_1, b_2\right) \\ var\left(b_1, b_2\right) & var\left(b_2\right) \end{bmatrix} = \begin{bmatrix} \frac{\left(\sum X_i\right)^2}{N\sum\left(X_i - \overline{X}\right)^2} & \frac{-\overline{X}}{\sum\left(X_i - \overline{X}\right)^2} \\ \frac{-\overline{X}}{\sum\left(X_i - \overline{X}\right)^2} & \frac{1}{\sum\left(X_i - \overline{X}\right)^2} \end{bmatrix} \tag{132}$$

1.3.7 Variance of forecast (advanced topic)

$$var\left(f\right)=var\left(\widehat{Y}_0\right)+var\left(\widehat{\varepsilon}_0\right) \tag{133}$$

$$var\left(\widehat{Y}_0\right)=\widehat{\sigma}^2\left[\frac{1}{N}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right] \tag{134}$$

$$var\left(f\right)=\widehat{\sigma}^2\left[\frac{1}{N}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right]+\widehat{\sigma}^2 \tag{135}$$

$$var\left(f\right)=\widehat{\sigma}^2\left[1+\frac{1}{N}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right] \tag{136}$$

1.3.8 Variance of prediction error

$$var\left(\widehat{\beta}_1\right)=\left[1+\frac{1}{N}+\frac{\left(x_0-\overline{x}\right)^2}{\sum\left(x_0-\overline{x}\right)_i^2}\right]\widehat{\sigma}^2 \tag{137}$$

Proof

$$Y_0=\widehat{Y}_0+\widehat{\varepsilon}_0 \tag{138}$$

$$var\left(Y_0\right)=var\left(\widehat{Y}_0\right)+var\left(\widehat{\varepsilon}_0\right) \tag{139}$$

$$var\left(\widehat{Y}_0\right)=var\left(\widehat{\beta}_1+\widehat{\beta}_2X_0\right)=var\left(\widehat{\beta}_1\right)+X_0^2var\left(\widehat{\beta}_2\right)+2X_0covar\left(\widehat{\beta}_1\widehat{\beta}_2\right) \tag{140}$$

1.3.9 Variance of prediction

$$\begin{aligned} var\left(\widehat{Y}_0\right) &= \frac{\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2+X_0^2\frac{\sum\left(X_i-\overline{X}\right)}{\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2 \\ &\quad +2X_0\left(-\overline{X}\frac{1}{\sum\left(X_i-\overline{X}\right)^2}\right)\widehat{\sigma}^2 \end{aligned} \tag{141}$$

add and subtract $\frac{N\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2$

$$\begin{aligned} var\left(\widehat{Y}_0\right) &= \frac{\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2-\frac{N\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2+X_0^2\frac{\sum\left(X_i-\overline{X}\right)}{\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2 \\ &\quad +2X_0\left(-\overline{X}\frac{1}{\sum\left(X_i-\overline{X}\right)^2}\right)\widehat{\sigma}^2+\frac{N\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\widehat{\sigma}^2 \end{aligned} \tag{142}$$

1.3.10 Variance of prediction

Taking common elements out

$$var\left(\widehat{Y}_0\right)=\widehat{\sigma}^2\left[\begin{array}{c}\frac{\sum\left(X_i-\overline{X}\right)^2-N\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}\\+\frac{X_0^2-2X_0\overline{X}+\sum\left(X_i-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\end{array}\right] \tag{143}$$

$$var\left(\widehat{Y}_0\right)=\widehat{\sigma}^2\left[\frac{\sum\left(X_i-\overline{X}\right)^2}{N\sum\left(X_i-\overline{X}\right)^2}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right] \tag{144}$$

1.3.11 Homework:

One major use of an econometric model is prediction. Suppose that a local supermarket wants you to estimate a model that determines expenditure on food in terms of income, and to predict food demand next year. Consider a simple regression model of the following form:

$$Y_t=\beta_1+\beta_2X_t+\varepsilon_t \qquad t=1...T \tag{145}$$

where Y_t is expenditure on food, X_t is income and ε_t is independently and identically distributed random error term with a zero mean and a constant variance.

From the sample information on food expenditure and income contained in "food.csv" file find estimated values of β_1 , β_2 and σ^2. You also want to predict the amount of expenditure on food Y_0 next year using information on likely income next year, X_0 . You may safely assume that as before $\varepsilon_0 \sim N(0,\sigma^2)$.

1. (a) Write down your prediction equation. Give an equation for the mean prediction, $E\left(Y_0\right)$.

   (b) What is your prediction of food expenditure if the income is £250? How can you compute your prediction error?

   (c) What is the variance of prediction error?

   (d) Construct a 95% confidence interval for your prediction. Explain what this interval means. How would you modify your model if the confidence interval of prediction is very large?

   (e) Give a graphical explanation of your answers in (a)-(d), labelling your diagrams carefully.

1.3.12 Exercise 2

A sport centre has a gym. A hypothetical data set on the monthly charges (X) and number of people using the gym (Y) are given in the following table with the values of cross products and square terms

Table 7: Monthy charges and number of customers

X_i	10	8	7	6	3	5	9	12	11	10
Y_i	60	75	90	100	150	120	125	100	80	65

1. (a) Represent X and Y in a Scattered diagram.
   (b) Draw horizontal and vertical lines with the mean of X and Y in that diagram.
   (c) Draw a line by your hand that best represents all sample observations.
   (d) Write a classical linear regression model in which X causes Y.
   (e) Write the assumptions of the error terms.
   (f) Derive normal equations of the OLS estimator minimising sum of squared errors. Estimate parameters of the model using above information. Use the deviation technique in your estimation.
   (g) What is your prediction of Y when X is 13?
   (h) Calculate the sum of variation in Y.
   (i) Decompose this total variance into explained and residual components.
   (j) Find the coefficient of determination or the R-square of this model.
   (k) Find the variance and standard error of the slope parameter.
   (l) Calculate the t-statistics and determine its level of significance using the T-table.
   (m) Construct a 95 percent confidence interval for the slope parameter.
   (n) Find the variance and the standard error of the intercept parameter.

1.4 Best, Linear, Unbiasedness of OLS estimators (BLUE)

The Gauss-Markov theorems for a linear regression model

Linearity of slope and intercept parameters:

- Consider a linear regression

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad i = 1 \ldots N \tag{146}$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \tag{147}$$

Intercept and slopes are linear on dependent varibales

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum w_i y_i \tag{148}$$

Where $w_i = \frac{x_i}{\sum x_i^2}$ is a constant.

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} = \frac{\sum y_i}{N} - \overline{X} \sum w_i y_i \tag{149}$$

- Thus $\widehat{\beta}_2$ and $\widehat{\beta}_1$ are linear on y_i

1.4.1 Unbiasedness of intercept parameter

$$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2\overline{X} \;=\; \frac{\sum y_i}{N} - \overline{X}\sum w_i y_i \tag{150}$$

$$E\left(\widehat{\beta}_1\right) = \; E\left(\frac{\sum\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right)}{N}\right) - E\left(\widehat{\beta}_2\overline{X}\;\right) \tag{151}$$

$$E\left(\widehat{\beta}_1\right) = \; E\left(\frac{N\beta_1}{N} + \frac{\beta_2\sum X_i}{N} + \frac{\sum\varepsilon_i}{N}\right) - E\left(\widehat{\beta}_2\overline{X}\;\right) \tag{152}$$

$$E\left(\widehat{\beta}_1\right) = \; \beta_1 + \beta_2\overline{X} - E\left(\widehat{\beta}_2\overline{X}\;\right) \tag{153}$$

$$\left(E\left(\widehat{\beta}_1\right) - \;\beta_1\right) = \overline{X}\left(\beta_2 - E\left(\widehat{\beta}_2\;\right)\right) \tag{154}$$

$$E\left(\widehat{\beta}_1\right) = \beta_1 \tag{155}$$

1.4.2 Unbiasedness of slope parameter

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum w_i y_i \tag{156}$$

$$E\left(\widehat{\beta}_2\right) = E\left(\sum w_i y_i\right) = E\sum w_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right) \tag{157}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_i\right) + \beta_2 E\left(\sum w_i x_i\right) + E\left(\sum w_i\varepsilon_i\;\right) \tag{158}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 \tag{159}$$

1.4.3 Minimum variance of slope parameter

$$E\left(\widehat{\beta}_2\right) = \sum w_i y_i \tag{160}$$

$$Var\left(\widehat{\beta}_2\right) = var\left[\frac{\sum\left(X_i - \overline{X}\right)}{\sum\left(X_i - \overline{X}\right)^2}\right] var\left(y_i\right) = \frac{1}{\sum x_i^2}\widehat{\sigma}^2 \tag{161}$$

Take b_2 any other linear and unbiased estimator. Then need to prove that $var(b_2) > var(\widehat{\beta}_2)$

$$E\left(b_2\right) = \sum k_i y_i \qquad k_i = w_i + c_i \tag{162}$$

$$E\left(b_2\right) \;=\; E\left[\sum k_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right)\right] \;= \tag{163}$$

$$E\left[\;\sum w_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right) + \sum c_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right)\right] \tag{164}$$

$$E\left(b_2\right)=E\left[\beta_1\sum w_i+\beta_2\sum w_ix_i+\sum w_i\varepsilon_i+\beta_1\sum c_i+\beta_2\sum c_ix_i+\sum c_i\varepsilon_i\ \right] \tag{165}$$

$$E\left(b_2\right)=\beta_2 \tag{166}$$

Minimum Variance of Slope Parameter (cont.)

$$E\left(b_2\right)=\beta_2 \tag{167}$$

$$var\left(b_2\right)=E\left[b_2-\beta_2\right]^2=E\left[\sum k_i\varepsilon_i\right]^2\ =E\left[\sum\left(w_i+c_i\right)\varepsilon_i\right]^2 \tag{168}$$

$$var\left(b_2\right)=\frac{1}{\sum x_i^2}\widehat{\sigma}^2+\widehat{\sigma}^2\sum c_i^2=var(\widehat{\beta}_2)+\widehat{\sigma}^2\sum c_i^2 \tag{169}$$

$$var(b_2)>var(\widehat{\beta}_2) \tag{170}$$

QED. Thus the OLS slope parameter is the best, linear and unbiased (BLUE). Similar proof can be applied for $Var\left(\widehat{\beta}_1\right)$.

Consistency of OLS Estimator: Large Sample or Assymptotic Property

$$Var\left(\widehat{\beta}_2\right)=\frac{1}{\sum x_i^2}\widehat{\sigma}^2 \tag{171}$$

$$\underset{\lim\quad N\rightarrow\infty}{Var\left(\widehat{\beta}_2\right)}\ =\underset{\lim\quad N\rightarrow\infty}{\frac{\frac{\widehat{\sigma}^2}{N}}{\frac{\sum x_i^2}{N}}}=0 \tag{172}$$

1.4.4 Covariance between slope and intercept parameters

$$\begin{aligned}
cov\left(\widehat{\beta}_1,\widehat{\beta}_2\right) &= E\left(\widehat{\beta}_1-E\left(\widehat{\beta}_1\right)\right)\left(\widehat{\beta}_2-E\left(\widehat{\beta}_2\right)\right)\\
&= E\left(\widehat{\beta}_1-\beta_1\right)\left(\widehat{\beta}_2-\beta_2\right)\\
&= -\overline{X}E\left(\widehat{\beta}_2-\beta_2\right)^2\\
&\because \left(E\left(\widehat{\beta}_1\right)-\ \beta_1\right)=\overline{X}\left(\beta_2-E\left(\widehat{\beta}_2\ \right)\right) && (173)\\
&= -\overline{X}\frac{1}{\sum x_i^2}\widehat{\sigma}^2 && (174)
\end{aligned}$$

.

1.4.5 Regression in matrix notations

Let Y is $N \times 1$ vector of dependent variables, X is $N \times K$ matrix of explanatory variables, e is $N \times 1$ vector of independently and identically distributed normal random variable with mean equal to zero and a constant variance $e \sim N(0, \sigma^2 I)$; β is a $K \times 1$ vector of unknown coefficients

$$Y = \beta X + e \tag{175}$$

Objective Objective is to choose β that minimise sum square errors

$$\begin{aligned} \underset{\beta}{Min}\; S(\beta) &= e'e = (Y - \beta X)'(Y - \beta X) \\ &= Y'Y - Y'(\beta X) - (\beta X)'Y + (\beta X)'(\beta X) && (176) \\ &= Y'Y - 2\beta X'Y + (\beta X)'(\beta X) && (177) \end{aligned}$$

First order condition:

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2\widehat{\beta}X'X = 0 \Longrightarrow \widehat{\beta} = (X'X)^{-1} X'Y \tag{178}$$

$$(X'X)^{-1} = \begin{pmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}^{-1} = \frac{1}{N\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & N \end{bmatrix} \tag{179}$$

$$(X'X)^{-1} = \begin{bmatrix} \frac{\sum X_i^2}{N\sum X_i^2 - (\sum X_i)^2} & -\frac{\sum X_i}{N\sum X_i^2 - (\sum X_i)^2} \\ -\frac{\sum X_i}{N\sum X_i^2 - (\sum X_i)^2} & \frac{N}{N\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \tag{180}$$

$$\widehat{\beta} = (X'X)^{-1} X'Y; \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2}{N\sum X_i^2 - (\sum X_i)^2} & -\frac{\sum X_i}{N\sum X_i^2 - (\sum X_i)^2} \\ -\frac{\sum X_i}{N\sum X_i^2 - (\sum X_i)^2} & \frac{N}{N\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \begin{bmatrix} \sum Y_i \\ \sum X_i' Y_i \end{bmatrix} \tag{181}$$

Derivation of Parameters (with Matrix Inverse)

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i' Y_i}{N\sum X_i^2 - (\sum X_i)^2} \\ \frac{N\sum X_i' Y_i - \sum X_i \sum Y_i}{N\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} = \begin{bmatrix} \frac{\sum X_i \sum X_i' Y_i - \sum X_i^2 \sum Y_i}{N\sum X_i^2 - (\sum X_i)^2} \\ \frac{\sum X_i \sum Y_i - N\sum X_i' Y_i}{N\sum X_i^2 - (\sum X_i)^2} \end{bmatrix} \tag{182}$$

Compare to what we had earlier:

$$\widehat{\beta}_2 = \frac{\sum X_i \sum Y_i - N\sum Y_i X_i}{(\sum X_i)^2 - N\sum X_i^2} = \frac{\sum x_i y_i}{\sum x_i^2} \tag{183}$$

$$\widehat{e} = \left(Y - \widehat{\beta} X\right) \tag{184}$$

$$\widehat{\sigma}^2 = \frac{\sum \widehat{e}_i^2}{N - k} = \frac{e'e}{N - k} \tag{185}$$

$$\sum \widehat{e}_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2 \tag{186}$$

$$\sum \widehat{y}^2 = \sum (x\beta)' (\beta x) \qquad x = X - \overline{X} \tag{187}$$

$$\sum \widehat{y}_i^2 = \sum (\widehat{\beta}_2 x_i)^2 = \widehat{\beta}_2^2 \sum {x_i}^2 \tag{188}$$

$$R^2 = \frac{\sum \widehat{y}_i^2}{\sum y_i^2} \text{ and } F_{calc} = \frac{\frac{\sum \widehat{y}_i^2}{K-1}}{\frac{\sum \widehat{e}_i^2}{N-K}};\ F_{calc} = \frac{R^2}{K-1}\frac{N-K}{(1-R^2)} \tag{189}$$

1.4.6 Variance in matrix notation

$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$

$$Y = \widehat{Y} + e = \widehat{\beta} X + e \tag{190}$$

$$Var(Y) = \sum y_i^2 = Y'Y - N\overline{Y}^2 \tag{191}$$

When there are two explantory variables in deviation from the mean form:

$$\widehat{y} = \widehat{\beta}_1 x_1 + \widehat{\beta}_1 x_2 \tag{192}$$

$$\begin{aligned}
\sum \widehat{y}^2 &= \left(\widehat{\beta}_1 \sum x_1 + \widehat{\beta}_2 \sum x_2\right)^2 = \widehat{\beta}_1^2 \sum x_1^2 + \widehat{\beta}_1 \widehat{\beta}_2 \sum x_1 x_2 + \widehat{\beta}_1 \widehat{\beta}_2 \sum x_1 x_2 + \widehat{\beta}_2^2 \sum x_2^2 \\
&= \widehat{\beta}_1 \left(\widehat{\beta}_1 \sum x_1^2 + \widehat{\beta}_2 \sum x_1 x_2\right) + \widehat{\beta}_2 \left(\widehat{\beta}_1 \sum x_1 x_2 + \widehat{\beta}_2 \sum x_1^2\right) \\
&= \widehat{\beta}_1 \sum x_1 y + \widehat{\beta}_2 \sum x_2 y
\end{aligned} \tag{193}$$

$$\sum \widehat{e}_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2 \tag{194}$$

$$\sum \widehat{y}^2 = \begin{bmatrix} \widehat{\beta}_1 & \widehat{\beta}_2 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & . & x_{1N} \\ x_{21} & x_{22} & . & x_{2N} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ . \\ y_N \end{bmatrix} = \widehat{\beta}' x' y \tag{195}$$

$$e'e = Y'Y - \widehat{\beta}' x' y \tag{196}$$

$$R^2 = \frac{\widehat{\beta}' x' y}{Y'Y} \text{ and } F_{calc} = \frac{\frac{\sum \widehat{y}_i^2}{K-1}}{\frac{e'e}{N-K}};\ F_{calc} = \frac{R^2}{K-1}\frac{N-K}{(1-R^2)} \tag{197}$$

1.4.7 Blue property in matrix: linearity and unbiasedness

$$\widehat{\beta} = (X'X)^{-1} X'Y \tag{198}$$

$$\widehat{\beta} = aY; \quad a = (X'X)^{-1} X' \tag{199}$$

Linearity proved.

$$E\left(\widehat{\beta}\right) = E\left[(X'X)^{-1} X' (X\beta + e)\right] \tag{200}$$

$$E\left(\widehat{\beta}\right) = E\left[(X'X)^{-1} X'X\beta\right] + E\left[(X'X)^{-1} X'e\right] \tag{201}$$

$$E\left(\widehat{\beta}\right) = \beta + E\left[(X'X)^{-1} X'e\right] \tag{202}$$

$$E\left(\widehat{\beta}\right) = \beta \tag{203}$$

Unbiasedness is proved.
Blue Property in Matrix: Minimum Variance

$$E\left(\widehat{\beta}\right) - \beta = E\left[(X'X)^{-1} X'e\right] \tag{204}$$

$$E\left[E\left(\widehat{\beta}\right) - \beta\right]^2 = E\left[(X'X)^{-1} X'e\right]'\left[(X'X)^{-1} X'e\right] \tag{205}$$

$$= (X'X)^{-1} X'XE(e'e)(X'X)^{-1} = \widehat{\sigma}^2 (X'X)^{-1} \tag{206}$$

Take an alternative estimator b

$$b = \left[(X'X)^{-1} X' + c\right] Y \tag{207}$$

$$b = \left[(X'X)^{-1} X' + c\right] (X\beta + e) \tag{208}$$

$$b - \beta = E\left[(X'X)^{-1} X'e + ce\right] \tag{209}$$

1.4.8 Blue property in matrix: minimum variance

Now it need to be shown that

$$cov(b) > cov\left(\widehat{\beta}\right) \tag{210}$$

Take an alternative estimator b

$$b - \beta = E\left[(X'X)^{-1} X'e + ce\right] \tag{211}$$

$$\begin{aligned} cov(b) &= E\left[(b-\beta)(b-\beta)'\right] \\ &= E\left[(X'X)^{-1}X'e+ce\right]\left[(X'X)^{-1}X'e+ce\right] \\ &= \sigma^2(X'X)^{-1}+\sigma^2c^2 \end{aligned} \tag{212}$$

$$cov(b) > cov\left(\widehat{\beta}\right) \tag{213}$$

Proved.
Thus the OLS is BLUE = Best, Linear, Unbiased Estimator.

1.5 Multiple Regression Model in Matrix

Consider a multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + ... + \beta_k X_{k,i} + \varepsilon_i \quad i = 1\ ...N \tag{214}$$

Assumptions:

$$E(\varepsilon_i) \; = 0 \tag{215}$$

$$E(\varepsilon_i x_{j,i}) \; = 0;\; var(\varepsilon_i) \; = \sigma^2 \quad for \;\; \forall \;\; i; \;\; \varepsilon_i \sim N\left(0,\sigma^2\right) \tag{216}$$

$$covar(\varepsilon_i \varepsilon_j) \; = 0 \tag{217}$$

Explanatory variables are uncorrelated.

$$E(X_{1,i}X_{1,j}) = 0 \tag{218}$$

Objective is to choose parameters that minimise the sum of squared errors

$$\underset{\widehat{\beta}_0\widehat{\beta}_1\widehat{\beta}_2...\widehat{\beta}_k}{Min\ S} = \sum \varepsilon_i^2 = \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1,i} - \widehat{\beta}_2 X_{2,i} - \widehat{\beta}_3 X_{3,i} - .... - \widehat{\beta}_k X_{k,i}\right)^2 \tag{219}$$

Derivation of Normal Equations

$$\frac{\partial S}{\partial \widehat{\beta}_0} = 0; \frac{\partial S}{\partial \widehat{\beta}_1} = 0; \frac{\partial S}{\partial \widehat{\beta}_2} = 0; \frac{\partial S}{\partial \widehat{\beta}_3} = 0; ...... \frac{\partial S}{\partial \widehat{\beta}_k} = 0 \tag{220}$$

1.5.1 Normal equations

Normal equations for two explanatory variable case

$$\sum Y_i = \widehat{\beta}_0 N + \widehat{\beta}_1 \sum X_{1,i} + \widehat{\beta}_2 \sum X_{2,i} \tag{221}$$

$$\sum X_{1,i}Y_i = \widehat{\beta}_0 \sum X_{1,i} + \widehat{\beta}_1 \sum X_{1,i}^2 + \widehat{\beta}_2 \sum X_{1,i}X_{2,i} \tag{222}$$

$$\sum X_{2,i}Y_i = \widehat{\beta}_0 \sum X_{2,i} + \widehat{\beta}_1 \sum X_{1,i}X_{2,i} + \widehat{\beta}_2 \sum X_{2,i}^2 \tag{223}$$

$$\begin{bmatrix} \sum Y_i \\ \sum X_{1,i}Y_i \\ \sum X_{2,i}Y_i \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \tag{224}$$

1.5.2 Normal equations in matrix form

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1,i} \\ \sum Y_i X_{2,i} \end{bmatrix} \tag{225}$$

$$\beta = (X'X)^{-1} X'Y \tag{226}$$

$$\widehat{\beta}_0 = \frac{\begin{vmatrix} \sum Y_i & \sum X_{1,i} & \sum X_{2,i} \\ \sum Y_i X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum Y_i X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{227}$$

1.5.3 Cramer Rule to solve for paramers

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} N & \sum Y_i & \sum X_{2,i} \\ \sum X_{1,i} & \sum Y_i X_{1,i} & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum Y_i X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{228}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} N & \sum X_{1,i} & \sum Y_i \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum Y_i X_{1,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum Y_i X_{2,i} \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{229}$$

Covariance of Parameters Matrix must be non-singular

$$(X'X)^{-1} \neq 0 \tag{230}$$

$$cov\left(\widehat{\beta}\right) = \begin{pmatrix} var(\widehat{\beta}_1) & cov(\widehat{\beta}_1\widehat{\beta}_2) & cov(\widehat{\beta}_1\widehat{\beta}_3) \\ cov(\widehat{\beta}_1\widehat{\beta}_2) & var(\widehat{\beta}_2) & cov(\widehat{\beta}_2\widehat{\beta}_3) \\ cov(\widehat{\beta}_1\widehat{\beta}_3) & cov(\widehat{\beta}_2\widehat{\beta}_3) & var(\widehat{\beta}_3) \end{pmatrix} \tag{231}$$

$$cov\left(\widehat{\beta}\right) = (X'X)^{-1} \sigma^2 \tag{232}$$

$$cov\left(\widehat{\beta}\right) = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{bmatrix}^{-1} \widehat{\sigma}^2 \tag{233}$$

Consider the deviation form:

Variance of errors
$\sum \widehat{e}_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2 = 23.12 - 21.64 = 1.48$

$$E\left(\widehat{\varepsilon}_i\right)^2 = \frac{\sum \widehat{e}_i^2}{N-k} = \widehat{\sigma}^2 \tag{234}$$

It is easier to consider normal equations in the deviation form:

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i x_{1,i} \\ \sum y_i x_{2,i} \end{bmatrix} \tag{235}$$

$$\beta = \left(X'X\right)^{-1} X'Y \tag{236}$$

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} \sum y_i x_{1,i} & \sum x_{1,i}x_{2,i} \\ \sum y_i x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}} \tag{237}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} \sum x_{1,i}^2 & \sum y_i x_{1,i} \\ \sum x_{1,i}x_{2,i} & \sum y_i x_{2,i} \end{vmatrix}}{\begin{vmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}} \tag{238}$$

$$\begin{aligned} & \begin{bmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix}^{-1} \\ = \quad & \frac{1}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2} \begin{bmatrix} \sum x_{2,i}^2 & -\sum x_{1,i}x_{2,i} \\ -\sum x_{1,i}x_{2,i} & \sum x_{1,i}^2 \end{bmatrix} \end{aligned} \tag{239}$$

$$var\left(\widehat{\beta}_1\right) = \frac{\sum x_{2,i}^2}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2}\sigma^2 \tag{240}$$

$$var\left(\widehat{\beta}_2\right) = \frac{\sum x_{1,i}^2}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2}\sigma^2 \tag{241}$$

1.5.4 Application

Numerical Example: Does level of unempolyment depend on claimant count, strikes and work hours?

How does the level of unemployment (Y_i)relate to the level of claimant counts ($X_{1,i}$), numbers of stopages($X_{2,i}$) because of industrial strikes and number of work hours ($X_{3,i}$)in UK? Data from the Labour Force Surve for 19 years;$N = 19$.

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i \quad i = 1 \ldots N \tag{242}$$

$$
\begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} & \sum X_{3,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{3,i} & \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix}
= \begin{bmatrix} 19 & 29057 & 4109 & 16904.6 \\ 29057 & 53709128.8 & 6872065.8 & 25461639.46 \\ 4109 & 6872065.8 & 1132419 & 3638145 \\ 16904.6 & 25461639.46 & 3638145 & 15059252.96 \end{bmatrix}
$$

Numerical Example: OLS Setup

$$
\begin{bmatrix} \sum Y_i \\ \sum X_{1,i}Y_i \\ \sum X_{2,i}Y_i \\ \sum X_{3,i}Y_i \end{bmatrix} = \begin{bmatrix} 37326 \\ 63415261.4 \\ 8476146 \\ 32958009 \end{bmatrix}
$$

$$
\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 19 & 29057 & 4109 & 16904.6 \\ 29057 & 53709128.8 & 6872065.8 & 25461639.46 \\ 4109 & 6872065.8 & 1132419 & 3638145 \\ 16904.6 & 25461639.46 & 3638145 & 15059252.96 \end{bmatrix}^{-1} \begin{bmatrix} 37326 \\ 63415261.4 \\ 8476146 \\ 32958009 \end{bmatrix}
$$

Numerical Example: Estimates of parameters, their standard errors and t-values

$$
\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 402.4319485 & -0.018888662 & 0.013959234 & -0.423181717 \\ -0.018888662 & 1.00E-06 & -9.85E-07 & 1.97E-05 \\ 0.013959234 & -9.85E-07 & 5.37E-06 & -1.53E-05 \\ -0.423181717 & 1.97E-05 & -1.53E-05 & 0.000445415 \end{bmatrix} \begin{bmatrix} 37326 \\ 63415261.4 \\ 8476146 \\ 32958009 \end{bmatrix}
$$

$$
\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} -5560.880967 \\ 0.98458293 \\ -0.223288328 \\ 6.820108368 \end{bmatrix}
$$

$$
\begin{bmatrix} SE\left(\widehat{\beta}_0\right) \\ SE\left(\widehat{\beta}_1\right) \\ SE\left(\widehat{\beta}_2\right) \\ SE\left(\widehat{\beta}_3\right) \end{bmatrix} = \begin{bmatrix} 871.2384101 \\ 0.04348893 \\ 0.100627026 \\ 0.916586018 \end{bmatrix}
$$

$$
\begin{bmatrix} t\left(\widehat{\beta}_0\right) \\ t\left(\widehat{\beta}_1\right) \\ t\left(\widehat{\beta}_2\right) \\ t\left(\widehat{\beta}_3\right) \end{bmatrix} = \begin{bmatrix} -6.382731641 \\ 22.63985164 \\ -2.218969753 \\ 7.440772858 \end{bmatrix}
$$

t-test

Hypotheses: $H_0 : \beta_i = 0$ against $H_A : \beta_i \neq 0$

Critical values of t for 15 degrees fo freedom at 5% level of significance = 2.13;

Each of above computed t-values are greater than table values. Therefore statistical enough evidence to reject the null hypothess. All of four parameters are statistically significant.

Numerical Example:Sum Square Error and Covariance of Beta

$$
var(e) = E\left(\widehat{\varepsilon}_i\right)^2 = \frac{\sum \widehat{e}_i^2}{N-k} = \frac{28292.59842}{19-4} = 1886.173228 = \widehat{\sigma}^2 \tag{243}
$$

$$
\begin{aligned}
cov\left(\widehat{\beta}\right) &= \begin{bmatrix} 402.4319485 & -0.018888662 & 0.013959234 & -0.423181717 \\ -0.018888662 & 1.00E-06 & -9.85E-07 & 1.97E-05 \\ 0.013959234 & -9.85E-07 & 5.37E-06 & -1.53E-05 \\ -0.423181717 & 1.97E-05 & -1.53E-05 & 0.000445415 \end{bmatrix} (1886.173228) \quad (244) \\
&= \begin{bmatrix} 759056.3673 & -35.62728928 & 26.32953287 & -798.1940248 \\ -35.62728928 & 0.001891287 & -0.001858782 & 0.037244366 \\ 26.32953287 & -0.001858782 & 0.010125798 & -0.028859446 \\ -798.1940248 & 0.037244366 & -0.028859446 & 0.840129929 \end{bmatrix}
\end{aligned}
$$

R-Square and Adjusted R-Square

$$
R^2 = \frac{4428800.138}{4457092.737} = 0.99365223 \tag{245}
$$

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{N-1}{N-K} = 0.992382676 \tag{246}$$

1.5.5 Analysis of Variance

Table 8: Testing overall significance by F-test

Source of Variance	Sum	Degrees of freedom	Mean	F-value
Total sum square (TSS)	4457092.737	18	247616.2632	
Regression Sum Square (RSS)	4428800.138	3	1476266.713	782.6782243
Sum of square error	28292.59842	15	1886.173228	

Hypothesis: $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$ or model is meaningless against $H_A : \beta_0 \neq \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$ or at least one $\beta_i \neq 0$ model explains something.

Critical values of F for degrees of freedom of 3 and 15 at 5 percent level of significance = 3.29.

Calculated F-statistics is much higher than critical value. Thereofre there is statistical evidence to reject the null hypothesis.

That means in general this model is statistically siginificant.

Table of results summarising all above calculations are presented as:

Table 9: Determinants of Unemployment Rate

	Coefficient	Stadard Error	t-value
Intercept	-5560.881	871.238	-6.383
Claiment Count	0.985	0.043	22.640
Stoppages	-0.223	0.101	-2.219
Workhours	6.820	0.917	7.441
$R^2 = 0.99$, $F = 782.7$, $N = 19$.			

Higher the number of claiment count higher the rate of unemployment - this makes intuitive sense. Greater the number of work-hours more people unemployed, this also is as expected as less extra jobs are created when people work more hours. European experience in the recent recession is good example where work sharing is common practice. Stoppages and industrial strikes reduce unemployment as such threats to employers reduce possibility of firing of existing workers.

Normal equations for K variables

$$\sum Y_i = \widehat{\beta}_0 N + \widehat{\beta}_1 \sum X_{1,i} + \widehat{\beta}_2 \sum X_{2,i} + \widehat{\beta}_3 \sum X_{3,i} + .. + \widehat{\beta}_k \sum X_{k,i} \tag{247}$$

$$\sum Y_i X_{1,i} = \widehat{\beta}_0 \sum X_{1,i} + \widehat{\beta}_1 \sum X_{1,i}^2 + \widehat{\beta}_2 \sum X_{1,i} X_{2,i} + . + \widehat{\beta}_k \sum X_{1,i} X_{k,i} \tag{248}$$

....

$$\sum Y_i X_{k,i} = \widehat{\beta}_0 \sum X_{k,i} + \widehat{\beta}_1 \sum X_{1,i} X_{k,i} + \widehat{\beta}_2 \sum X_{k,i} X_{2,i} + . + \widehat{\beta}_k \sum X_{k,i}^2 \tag{249}$$

Process is similar to the three variable model - except that this general model will have more coefficients to evaluate and test; and requires data on more variables.

1.5.6 Exercise 3

Suppose you have the following data set on number of tickets sold in a football match (Y), price of tickets (X_1) and income of the customers (X_2). and Y are measured in 10 thousand pounds. You want to find out the exact relation between tickets sold and prices and income of people watching football games.

Table 10: Price, Income and Sales

$X_{1,i}$	11	7	6	5	3	2	1
$X_{2,i}$	2	2	4	5	6	5	4
Y_i	1	2	3	4	5	6	7

1. (a) Write a simple regression model to explain the number of tickets sold in terms of the price of the ticket. Explain briefly underlying assumptions and expected signs of the parameters in this model.
   (b) Estimate the slope and intercept parameters. Calculate cross products and squared terms needed for estimation from the above data table.
   (c) Use your estimates in (b) find the explained squared sum $\sum \widehat{y}_i^2$, sum of squared errors $\sum \widehat{e}_i^2$ and the $R^2$and $\overline{R}^2$.
   (d) Estimate the variance of the error term and the slope coefficient. Explain its importance.
   (e) Test whether the slope term is significant at 5% confidence level.
   (f) Build 95 percent confidence interval for estimate of slope and intercept terms.
   (g) Discuss how reducing type I error may cause increase in type II errors.
   (h) Calculate the elasticity of demand for football around the mean of Y and X_1.
   (i) Write a multiple regression model to explain the number of tickets sold in terms of the price of the ticket and the income of individuals going to the football game. What additional assumption(s) do you need while introducing an additional variable.
   (j) Estimate the parameters of that multiple regression model.
   (k) What is your prediction of the number of tickets sold if X_1= 5 and X_2=4?
   (l) Introduce dummy variables in your multiple regression model to show differences in demand for football ticket based on gender differences (1 for male and 0 for females), four seasons (autumn, winter, spring and summer) and interaction between gender and income.

1.5.7 Homework

Suppose that a leading supermarket in the city centre requests to estimate a demand function for beef. Your are considering estimating a model where demand for beef depends on price of beef , price of pork, price of chicken and consumer income as following:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \beta_4 X_{4,i} + \varepsilon_i \quad i = 1 ...N \tag{250}$$

where is Y_i demand for beef, $X_{1,i}$ is price of beef, $X_{2,i}$ is price of pork, $X_{3,i}$ is price of chicken,$X_{4,i}$ is income of consumer and ε_i is a normally and $\varepsilon_i \sim N(0, \sigma^2)$ identically distributed random variable.

1. (a) Using your knowledge of microeconomics, write down the expected signs of β_0, β_1, β_2, β_3,and β_4 in this model and explain why?
   (b) Write major assumptions of the ordinary least square approach to this model.
   (c) Suppose you have a data set on these variables over last 35 years and you want to estimate parameters β_0, β_1, β_2, β_3,and β_4. Derive normal equations that you will use get OLS estimators of these parameters?
   (d) Compute the variances of parameters β_1, β_2, β_3,and β_4.
   (e) Compute variance-covariance matrix for the random term.
   (f) Construct a confidence interval on β_1, β_2, β_3, and β_4 and predicted Y_i.
   (g) How would your result be affected if you find that $X_{1,i} = 0.6X_{2,i}$?
   (h) How would you modify your model to correct a problem in reported in (g)?

1.5.8 Testing for Restrictions

Multiple Regression Model in Matrix

- Consider a linear regression

$$Y_i = \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i \quad i = 1 \ldots N \tag{251}$$

and assumptions

$$E(\varepsilon_i) = 0 \tag{252}$$
$$E(\varepsilon_i x_{j,i}) = 0 \tag{253}$$
$$var(\varepsilon_i) = \sigma^2 \quad for \ \forall \ i \tag{254}$$
$$covar(\varepsilon_i \varepsilon_j) = 0 \tag{255}$$
$$\varepsilon_i \tilde{} N\left(0, \sigma^2\right) \tag{256}$$

- Objective is to choose parameters that minimise the sum of squared errors

$$\underset{\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3}{Min} S = \sum \varepsilon_i^2 = \left(Y_i - \widehat{\beta}_1 X_{1,i} - \widehat{\beta}_2 X_{2,i} - \widehat{\beta}_3 X_{3,i}\right)^2 \tag{257}$$

Derivation of Normal Equations

$$\frac{\partial S}{\partial \widehat{\beta}_1} = 0; \frac{\partial S}{\partial \widehat{\beta}_2} = 0; \frac{\partial S}{\partial \widehat{\beta}_3} = 0; \tag{258}$$

- Normal equations for three explanatory variable case

$$\sum X_{1,i}Y_i = \widehat{\beta}_1 \sum X_{1,i}^2 + \widehat{\beta}_2 \sum X_{1,i}X_{2,i} + \widehat{\beta}_3 \sum X_{1,i}X_{3,i} \tag{259}$$

$$\sum X_{2,i}Y_i = \widehat{\beta}_1 \sum X_{1,i}X_{2,i} + \widehat{\beta}_2 \sum X_{2,i}^2 + \widehat{\beta}_3 \sum X_{2,i}X_{3,i} \tag{260}$$

$$\sum X_{3,i}Y_i = \widehat{\beta}_1 \sum X_{1,i}X_{3,i} + \widehat{\beta}_2 \sum X_{2,i}X_{3,i} + \widehat{\beta}_3 \sum X_{3,i}^2 \tag{261}$$

$$\begin{bmatrix} \sum X_{1,i}Y_i \\ \sum X_{2,i}Y_i \\ \sum X_{3,i}Y_i \end{bmatrix} = \begin{bmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} \tag{262}$$

Normal equations in matrix form

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum X_{1,i}Y_i \\ \sum X_{2,i}Y_i \\ \sum X_{3,i}Y_i \end{bmatrix} \tag{263}$$

$$\beta = \left(X'X\right)^{-1} X'Y \tag{264}$$

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} \sum X_{1,i}Y_i & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{2,i}Y_i & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{3,i}Y_i & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{vmatrix}}{\begin{vmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{vmatrix}} \tag{265}$$

Use Cramer Rule to solve for paramers

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}Y_i \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}Y_i \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}Y_i \end{vmatrix}}{\begin{vmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{vmatrix}} \tag{266}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} \sum X_{1,i}^2 & \sum X_{1,i}Y_i & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}Y_i & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{3,i}Y_i & \sum X_{3,i}^2 \end{vmatrix}}{\begin{vmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{vmatrix}} \tag{267}$$

Covariance of Parameters

1.5.9 Matrix must be non-singular

$$(X'X)^{-1} \neq 0 \tag{268}$$

$$cov\left(\widehat{\beta}\right) = \begin{pmatrix} var(\widehat{\beta}_1) & var(\widehat{\beta}_1\widehat{\beta}_2) & var(\widehat{\beta}_1\widehat{\beta}_3) \\ var(\widehat{\beta}_1\widehat{\beta}_2) & var(\widehat{\beta}_2) & var(\widehat{\beta}_2\widehat{\beta}_3) \\ var(\widehat{\beta}_1\widehat{\beta}_3) & var(\widehat{\beta}_2\widehat{\beta}_3) & var(\widehat{\beta}_3) \end{pmatrix} \tag{269}$$

$$cov\left(\widehat{\beta}\right) = (X'X)^{-1}\sigma^2 \tag{270}$$

$$cov\left(\widehat{\beta}\right) = \begin{bmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix}^{-1} \widehat{\sigma}^2 \tag{271}$$

Data (text book example Carter, Griffith and Hill)

Table 11: Data for a multiple regression

y	1	-1	2	0	4	2	2	0	2
x1	1	-1	1	0	1	0	0	1	0
x2	0	1	0	1	2	3	0	-1	0
x3	-1	0	0	0	0	0	1	1	1

$$\begin{bmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 16 & -1 \\ 0 & -1 & 4 \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} \sum X_{1,i}Y_i \\ \sum X_{2,i}Y_i \\ \sum X_{3,i}Y_i \end{bmatrix} = \begin{bmatrix} 8 \\ 13 \\ 3 \end{bmatrix}$$

Estimation of Parameters

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 16 & -1 \\ 0 & -1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 8 \\ 13 \\ 3 \end{bmatrix} \tag{272}$$

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.063 & 0.016 \\ 0 & 0.016 & 0.254 \end{bmatrix} \begin{bmatrix} 8 \\ 13 \\ 3 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.873 \\ 0.968 \end{bmatrix} \tag{273}$$

Estimated equation

$$\widehat{Y}_i = 1.6X_{1,i} + 0.873X_{2,i} + 0.968X_{3,i} \tag{274}$$

Estimation of Errors

$$\widehat{e}_i = Y_i - 1.6X_{1,i} + 0.873X_{2,i} + 0.968X_{3,i} \tag{275}$$

$$\widehat{e}_1 = 1 - 1.6\,(1) + 0.873\,(0) + 0.968\,(-1) = 0.368 \tag{276}$$

$$\widehat{e}_2 = -1 - 1.6\,(-1) + 0.873\,(1) + 0.968\,(0) = -0.273 \tag{277}$$

$$\widehat{e}_3 = 2 - 1.6\,(1) + 0.873\,(0) + 0.968\,(0) = 0.4 \tag{278}$$

$$\widehat{e}_4 = 0 - 1.6\,(0) + 0.873\,(1) + 0.968\,(0) = -0.873 \tag{279}$$

$$\widehat{e}_5 = 4 - 1.6\,(1) + 0.873\,(2) + 0.968\,(0) = 0.654 \tag{280}$$

$$\widehat{e}_6 = 2 - 1.6\,(0) + 0.873\,(3) + 0.968\,(0) = -0.619 \tag{281}$$

$$\widehat{e}_7 = 2 - 1.6\,(0) + 0.873\,(0) + 0.968\,(1) = 1.032 \tag{282}$$

$$\widehat{e}_8 = 0 - 1.6\,(1) + 0.873\,(-1) + 0.968\,(1) = -1.695 \tag{283}$$

$$\widehat{e}_9 = 2 - 1.6\,(0) + 0.873\,(0) + 0.968\,(1) = 1.032 \tag{284}$$

Sum of Error square, variance and covariance of Beta

$$\begin{aligned}
\sum \widehat{e}_i^2 &= 0.368^2 + (-0.273)^2 + 0.4^2 \\
&\quad + (-0.873)^2 + (0.654)^2 + (-0.619)^2 + 1.032^2 \\
+ (-1.695)^2 + 1.032^2 &= 6.9460
\end{aligned} \tag{285}$$

Variance of errors

$$var(e) = E\left(\widehat{\varepsilon}_i\right)^2 = \frac{\sum \widehat{e}_i^2}{N-k} = \frac{6.9460}{9-3} = 1.1577 = \widehat{\sigma}^2 \tag{286}$$

$$\begin{aligned}
cov\left(\widehat{\beta}\right) &= \begin{bmatrix} \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i}X_{3,i} \\ \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i}X_{3,i} \\ \sum X_{1,i}X_{3,i} & \sum X_{2,i}X_{3,i} & \sum X_{3,i}^2 \end{bmatrix}^{-1} \widehat{\sigma}^2 \\
&= \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.063 & 0.016 \\ 0 & 0.016 & 0.254 \end{bmatrix} (1.1577) = \begin{bmatrix} 0.232 & 0 & 0 \\ 0 & 0.074 & 0.018 \\ 0 & 0.018 & 0.294 \end{bmatrix}
\end{aligned} \tag{287}$$

Test of Restrictions

$$var\left(\widehat{\beta}_1\right) = 0.232; var\left(\widehat{\beta}_2\right) = 0.074; var\left(\widehat{\beta}_1\right) = 0.294; \tag{288}$$

$$cov\left(\widehat{\beta}_1\widehat{\beta}_2\right) = cov\left(\widehat{\beta}_1\widehat{\beta}_3\right) = 0; cov\left(\widehat{\beta}_2\widehat{\beta}_3\right) = cov\left(\widehat{\beta}_3\widehat{\beta}_2\right) = 0; \tag{289}$$

F-test

$$F = \frac{(Rb - r)' \left[Rcov\left(b\right) R'\right]^{-1} (Rb - r)}{J} \tag{290}$$

Hypothesis H_0: $\beta_1 = \beta_2 = \beta_3 = 0$ against H_A: $\beta_1 \neq 0$; $\beta_2 \neq 0$; or $\beta_3 \neq 0$
Here $J = 3$ is the number of restrictions

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; b = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix}; r = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{291}$$

Test of Restrictions

$$F = \frac{\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right)' \left[\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0.232 & 0 & 0 \\ 0 & 0.074 & 0.018 \\ 0 & 0.018 & 0.294 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}'\right]^{-1} \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right)}{J = 3} \tag{292}$$

See matrix_restrictions.xls for calculations.
Test of Restrictions

$$F = \frac{\begin{pmatrix} 1.6 & 0.873 & 0.968 \end{pmatrix}\begin{bmatrix} 4.3190 & 0 & 0 \\ 0 & 13.821 & -0.8638 \\ 0 & -0.8638 & 3.455 \end{bmatrix}\begin{bmatrix} 1.6 \\ 0.873 \\ 0.968 \end{bmatrix}}{3} \tag{293}$$

$$F = \frac{\begin{pmatrix} 1.6 & 0.873 & 0.968 \end{pmatrix}\begin{bmatrix} 6.91042 \\ 11.22943 \\ 2.59141 \end{bmatrix}}{3} = \frac{23.37}{3} = 7.79 \tag{294}$$

$F_{(m1,m2),\alpha} = F_{(3,6),5\%} = 4.76$;
Critical value for F at degrees of freedom of (3,6) at 5% confidence interval is 4.76.
F calculated is bigger than F critical => Reject null hypothesis, which says

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

At least one of these parameters is significant and explains variation in y, in other words accept

$$H_A\text{: } \beta_1 \neq 0;\ \ \beta_2 \neq 0; \text{ or } \beta_3 \neq 0$$

1.5.10 Dummy variables in a regression model

- It represents qualitative aspect or characteristic in the data
  Quality : good, bad; Location: south/north/east/west; characterisitcs: fat/thin or tall/short
  Time: Annual 1970s/ 1990s.; seasonal: Summer, Autumn, Winter, Spring;
- Gender: male/female; Education: GCSE/UG/PD/PhD
  Subjects: Math/English/Science/Economics
- Ethnic backgrounds: Black, White, Asian, Cacasian, European, American, Latinos, Mangols, Ausis.

$$Y_i = \beta_1 + \beta_2 X_i + \beta_2 D_i + \varepsilon_i \quad i = 1 \text{ ...} N \tag{295}$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right) \tag{296}$$

- Here D_i is special type of variable

$$D_i = \int \begin{array}{l} 1 = \text{if the certain quality exists} \\ 0 = \text{otherwise} \end{array} \tag{297}$$

Dummy Variables in a Regression Model

- Three types of dummy
  1. Slope dummy
  2. Intercept dummy
  3. Interaction between slope and intercept
     Examples
     - Earnding differences by gender, region, ethnicity or religion, occupation, education level.
     - Unemployment duration by gender, region, ethnicity or religion, occupation, education level.
     - Demand for a product by by weather, season, gender, region, ethnicity or religion, occupation, education level.
     - Test scores by gender, previous background, ethnic origin
     - Growth rates by decades, countries, exchange rate regimes

Dummy Variables Trap: Consider seasonal dummies as

$$Y_i = \beta_1 + \beta_2 X_i + \beta_2 D_1 + \beta_2 D_2 + \beta_2 D_3 + \beta_2 D_4 + \varepsilon_i \tag{298}$$

where

$$D_1 = \int \begin{array}{l} 1 = \text{if summer} \\ 0 \text{ otherwise} \end{array} \tag{299}$$

$$D_2 = \int \begin{matrix} 1 = \text{if autumn} \\ 0 \text{ otherwise} \end{matrix} \tag{300}$$

$$D_3 = \int \begin{matrix} 1 = \text{if winter} \\ 0 \text{ otherwise} \end{matrix} \tag{301}$$

$$D_4 = \int \begin{matrix} 1 = \text{if spring} \\ 0 \text{ otherwise} \end{matrix} \tag{302}$$

- Since $\sum D_i = 1$, it will cause multicollinearity as:

$$D_1 + D_2 + D_3 + D_4 = 1 \tag{303}$$

  drop on of D_i to avoid the dummy variable trap.

Dummy Variables in a piecewise linear regression models

- Threshold effects in sales
- tariff charges by volume of transaction -mobile phones
- Panel regression: time and individual dummies
- Pay according to hierarchy in an organisation
- profit from whole sale and retail sales
- age dependent earnings -Scholarship for students, pensions and allowances for elderly
- tax allowances by level of income or business
- Investment credit by size of investment
- prices, employeemnts, profits or sales for small, medium and large scale corporations
- requirements according to weight or hight of body

1.5.11 Test of structural change

Chow Test for stability of parameters or structural change

- Use n1 and n2 observations to estimate overall and separate regressions with (n1+n2-k, n1-k, and n2-k) degrees of freedoms;
- obtain SSR1(with n1+n2-k dfs),
- SSR2 (with n1-k dfs),
- SSR3 (with n2-k dfs) and
- SSR4 = SSR1+ SSR2 (with n1+n2-2k dfs),
- obtain S5 = S1-S4;

- do F-test

$$F = \frac{\frac{S_5}{k}}{\frac{S_5}{(n1+n2-2k)}} \tag{304}$$

The advantage of this approach to the Chow test is that it does not require the construction of the dummy and interaction variables.

1.6 Exercise 4

Suppose that you are interested in estimating the demand for beer in Yorkshire pubs and consider the following multiple regression model:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_{1,i}) + \beta_2 \ln(X_{2,i}) + \beta_3 \ln(X_{3,i}) + \beta_4 \ln(X_{4,i}) + \varepsilon_i \quad i = 1 \ldots N \tag{305}$$

where Y_i is the demand for beer, $X_{1,i}$ is the price of beer, $X_{2,i}$ is the price of other liquor products, $X_{3,i}$ is the price of food and other services, $X_{4,i}$ is consumer income. Coefficients β_0,β_1,β_2,β_3,and β_4 are the set of unknown elasticity coefficients you would like to estimate. Again assume that errors ε_i are independently normally distributed, $\varepsilon_i \sim N(0, \sigma^2)$.

1. (a) Estimate the unknown parameters of this model using data in Beer1.csv.
   (b) How would you determine the overall significance of this model? Write down your test criterion. Compare that test statistic with another test statistic that you would use to test whether a particular coefficient, such as β_3, is statistically significant or not.
   (c) How would you establish whether a particular variable is helping to explain the variation in beer consumption?
   (d) Further suppose that you have some non-sample information on the relation between the price and income coefficients as following:
      i. sum of the elasticities equals zero: β_1+β_2+β_3+ $\beta_4 = 0$.
      ii. two cross elasticities are equal: β_3=$\beta_4 = 0$ or β_3- $\beta_4 = 0$
      iii. income elasticity is equal to unity: $\beta_5 = 1$
   (e) How do you test whether these restrictions are valid or not ?
   (f) In addition to the variables listed in the above model you suspect that gender and level of education of individuals are important determinants of beer consumption. Explain how you could incorporate these variables in this model.
   (g) The income of an individual also depends upon his/her age. Income in turn determines the consumption of beer. Thus age interacts with income. How would you introduce this age-income interaction effect in the above model?

Instructions for testing linear restrictions in PcGive for cross section data like this:
a. regress Y on $X_{1,i}$ $X_{2,i}$, $X_{3,i}$ and $X_{4,i}$.
b. click on test/linear restriction, put the restrictions in the matrix box. one line for each restriction. For instance if β_0+β_1+β_2+β_3+ $\beta_4 = 0$. to be tested then type 1 1 1 1 1 0 , then click ok , it will test validity of that restriction. If there are two restriction

β_0+β_1+β_2+β_3+ $\beta_4 = 0$ and β_3- $\beta_4 = 0$ then
1 1 1 1 1 0
0 0 0 1 -1 0
put this input in the matrix box, then click OK. This will test for thoth restrictions.

1.7 Multicollinearity

Multiple Regression Model in Matrix

- Consider a linear regression

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + ... + \beta_k X_{k,i} + \varepsilon_i \quad i = 1 \; ...N \tag{306}$$

and assumptions

$$E\left(\varepsilon_i\right) \;\; = 0 \tag{307}$$

$$E\left(\varepsilon_i x_{j,i}\right) \;\; = 0; \; var\left(\varepsilon_i\right) \;\; = \sigma^2 \quad for \;\; \forall \;\; i; \;\; \varepsilon_i \sim N\left(0, \sigma^2\right) \tag{308}$$

$$covar\left(\varepsilon_i \varepsilon_j\right) \;\; = 0 \tag{309}$$

Explanatory variables are uncorrelated.

$$E\left(X_{1,i} X_{1,j}\right) = 0 \tag{310}$$

- Objective is to choose parameters that minimise the sum of squared errors

$$\underset{\widehat{\beta}_0 \widehat{\beta}_1 \widehat{\beta}_2 ... \widehat{\beta}_k}{Min \; S} = \sum \varepsilon_i = \left(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1,i} - \widehat{\beta}_2 X_{2,i} - \widehat{\beta}_3 X_{3,i} - .... - \widehat{\beta}_k X_{k,i}\right) \tag{311}$$

Derivation of Normal Equations

$$\frac{\partial S}{\partial \widehat{\beta}_0} = 0; \frac{\partial S}{\partial \widehat{\beta}_1} = 0; \frac{\partial S}{\partial \widehat{\beta}_2} = 0; \frac{\partial S}{\partial \widehat{\beta}_3} = 0; ......\frac{\partial S}{\partial \widehat{\beta}_k} = 0 \tag{312}$$

- Normal equations for two explanatory variable case

$$\sum Y_i = \widehat{\beta}_0 N + \widehat{\beta}_1 \sum X_{1,i} + \widehat{\beta}_2 \sum X_{2,i} \tag{313}$$

$$\sum X_{1,i} Y_i = \widehat{\beta}_0 \sum X_{1,i} + \widehat{\beta}_1 \sum X_{1,i}^2 + \widehat{\beta}_2 \sum X_{1,i} X_{2,i} \tag{314}$$

$$\sum X_{2,i} Y_i = \widehat{\beta}_0 \sum X_{2,i} + \widehat{\beta}_1 \sum X_{1,i} X_{2,i} + \widehat{\beta}_2 \sum X_{2,i}^2 \tag{315}$$

$$\begin{bmatrix} \sum Y_i \\ \sum X_{1,i} Y_i \\ \sum X_{2,i} Y_i \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i} X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i} X_{2,i} & \sum X_{2,i}^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \tag{316}$$

Normal equations in matrix form

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1,i} \\ \sum Y_i X_{2,i} \end{bmatrix} \tag{317}$$

$$\beta = \left(X'X\right)^{-1} X'Y \tag{318}$$

$$\widehat{\beta}_0 = \frac{\begin{vmatrix} \sum Y_i & \sum X_{1,i} & \sum X_{2,i} \\ \sum Y_i X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum Y_i X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{319}$$

Use Cramer Rule to solve for paramers

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} N & \sum Y_i & \sum X_{2,i} \\ \sum X_{1,i} & \sum Y_i X_{1,i} & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum Y_i X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{320}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} N & \sum X_{1,i} & \sum Y_i \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum Y_i X_{1,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum Y_i X_{2,i} \end{vmatrix}}{\begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}} \tag{321}$$

Evaluate the determinant

$$|X'X| = \begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix} \tag{322}$$

$$|X'X| = N \sum X_{1,i}^2 \sum X_{2,i}^2 + \sum X_{1,i} \sum X_{1,i}X_{2,i} \sum X_{2,i} + \sum X_{2,i} \sum X_{1,i}X_{2,i} \sum X_{1,i}$$
$$- \sum X_{2,i} \sum X_{2,i} \sum X_{1,i}^2 - N \sum X_{1,i}X_{2,i} \sum X_{1,i}X_{2,i} - \sum X_{2,i}^2 \sum X_{1,i} \sum X_{1,i}$$

For this calculation, repeate first two columns as

$$|X'X| = \begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} & N & \sum X_{1,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} & \sum X_{1,i} & \sum X_{1,i}^2 \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 & \sum X_{2,i} & \sum X_{1,i}X_{2,i} \end{vmatrix} \tag{323}$$

Dederminant = (sume of cross product from to left to right - sum of cross product from bottom left to right)

1.7.1 Exact pulticollinearity: singularity

Significant R^2 but insignificant t-ratios. why?

In existence of exact multicollinearity $X'X$ is singular, i.e. $|X'X| = 0$

$X_{1,i} = \lambda X_{2,i}$

$|X'X| = N\sum X_{1,i}^2 \sum X_{2,i}^2 + \sum X_{1,i} \sum X_{1,i}X_{2,i} \sum X_{2,i} + \sum X_{2,i} \sum X_{1,i}X_{2,i} \sum X_{1,i}$
$- \sum X_{2,i} \sum X_{2,i} \sum X_{1,i}^2 - N\sum X_{1,i}X_{2,i} \sum X_{1,i}X_{2,i} - \sum X_{2,i}^2 \sum X_{1,i} \sum X_{1,i}$

Substituting out $X_{1,i}$

$|X'X| = N\lambda^2 \sum X_{2,i}^2 \sum X_{2,i}^2 + \lambda^2 \sum X_{2,i} \sum X_{2,i}^2 \sum X_{2,i} + \lambda^2 \sum X_{2,i} \sum X_{2,i}^2 \sum X_{2,i}$
$-\lambda^2 \sum X_{2,i} \sum X_{2,i} \sum X_{2,i}^2 - N\lambda^2 \sum X_{2,i}^2 \sum X_{2,i}^2 - \lambda^2 \sum X_{2,i}^2 \sum X_{2,i} \sum X_{2,i}$
$= 0$

$$|X'X| = \begin{vmatrix} N & \lambda\sum X_{2,i} & \sum X_{2,i} \\ \lambda\sum X_{2,i} & \lambda^2\sum X_{2,i}^2 & \lambda\sum X_{2,i}X_{2,i} \\ \sum X_{2,i} & \lambda\sum X_{2,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix} = 0 \tag{324}$$

Parameters are indeterminate in model with exact multicollinearity

$$\widehat{\beta}_0 = \frac{\begin{vmatrix} \sum Y_i & \sum X_{1,i} & \sum X_{2,i} \\ \sum Y_i X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum Y_i X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{0} = \infty \tag{325}$$

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} N & \sum Y_i & \sum X_{2,i} \\ \sum X_{1,i} & \sum Y_i X_{1,i} & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum Y_i X_{2,i} & \sum X_{2,i}^2 \end{vmatrix}}{0} = \infty \tag{326}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} N & \sum X_{1,i} & \sum Y_i \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum Y_i X_{1,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum Y_i X_{2,i} \end{vmatrix}}{0} = \infty \tag{327}$$

Covariance of parameters cannot be estimated in model with exact multicollinearity

$$(X'X)^{-1} = \infty \tag{328}$$

$$cov\left(\widehat{\beta}\right) = \begin{pmatrix} var(\widehat{\beta}_1) & cov(\widehat{\beta}_1\widehat{\beta}_2) & cov(\widehat{\beta}_1\widehat{\beta}_3) \\ cov(\widehat{\beta}_1\widehat{\beta}_2) & var(\widehat{\beta}_2) & cov(\widehat{\beta}_2\widehat{\beta}_3) \\ cov(\widehat{\beta}_1\widehat{\beta}_3) & cov(\widehat{\beta}_2\widehat{\beta}_3) & var(\widehat{\beta}_3) \end{pmatrix} = \infty \tag{329}$$

$$cov\left(\widehat{\beta}\right) = (X'X)^{-1}\sigma^2 = \infty \tag{330}$$

$$cov\left(\widehat{\beta}\right) = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{bmatrix}^{-1} \widehat{\sigma}^2 = \infty \tag{331}$$

Numerical example of exact multicollinearity

Table 12: Data for testing multicollinearity

y	3	5	7	6	9	6	7
x1	1	2	3	4	5	6	7
x2	5	10	15	20	25	30	35

Evaluate the determinant

$$|X'X| = \begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i}X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i}X_{2,i} & \sum X_{2,i}^2 \end{vmatrix} = \begin{vmatrix} 7 & 28 & 140 \\ 28 & 140 & 700 \\ 140 & 700 & 3500 \end{vmatrix}; \tag{332}$$

$$\begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1,i} \\ \sum Y_i X_{2,i} \end{bmatrix} = \begin{bmatrix} 43 \\ 188 \\ 980 \end{bmatrix}$$

Numerical example of exact multicollinearity

$|X'X| = N \sum X_{1,i}^2 \sum X_{2,i}^2 + \sum X_{1,i} \sum X_{1,i}X_{2,i} \sum X_{2,i} + \sum X_{2,i} \sum X_{1,i}X_{2,i} \sum X_{1,i}$
$- \sum X_{2,i} \sum X_{2,i} \sum X_{1,i}^2 - N \sum X_{1,i}X_{2,i} \sum X_{1,i}X_{2,i} - \sum X_{2,i}^2 \sum X_{1,i} \sum X_{1,i}$
$= (7 \times 140 \times 3500 + 28 \times 700 \times 140 + 140 \times 700 \times 28$
$-140 \times 140 \times 140 - 7 \times 700 \times 700 - 28 \times 28 \times 3500) = 0$

You can evaluate determinants easily in excel using following steps:

1. select the cell where to put the result.and press shift and control continously by two fingers of left hand
2. use mouse by right hand to choose math and trig function
3. choose MDETERM
4. Select matrix for which to evaluate the determinant
5. press OK and you will see the reslut.

Normal equations of a multiple regression in deviation form:

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i x_{1,i} \\ \sum y_i x_{2,i} \end{bmatrix} \tag{333}$$

$$\beta = (X'X)^{-1} X'Y \tag{334}$$

$$\widehat{\beta}_1 = \frac{\begin{vmatrix} \sum y_i x_{1,i} & \sum x_{1,i}x_{2,i} \\ \sum y_i x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}}{\begin{vmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}} \tag{335}$$

$$\widehat{\beta}_2 = \frac{\begin{vmatrix} \sum x_{1,i}^2 & \sum y_i x_{1,i} \\ \sum x_{1,i}x_{2,i} & \sum y_i x_{2,i} \end{vmatrix}}{\begin{vmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{vmatrix}} \tag{336}$$

Variances of parameters:

$$\begin{aligned}
&\begin{bmatrix} \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix}^{-1} \\
= &\frac{1}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2} \begin{bmatrix} \sum x_{2,i}^2 & -\sum x_{1,i}x_{2,i} \\ -\sum x_{1,i}x_{2,i} & \sum x_{1,i}^2 \end{bmatrix} \qquad (337)
\end{aligned}$$

$$var\left(\widehat{\beta}_1\right) = \frac{\sum x_{2,i}^2}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2}\sigma^2 \qquad (338)$$

$$var\left(\widehat{\beta}_2\right) = \frac{\sum x_{1,i}^2}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2}\sigma^2 \qquad (339)$$

Variance Inflation Factor in Inexact Multicollinearity
Let correlations between $X_{1,i}$ and $X_{2,i}$ be given by r_{12}. Then Variance inflation factor is $\frac{1}{(1-r_{12}^2)}$

$$\begin{aligned}
var\left(\widehat{\beta}_2\right) &= \frac{\sum x_{1,i}^2}{\left[\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2\right]}\sigma^2 \\
&= \frac{1}{\left[\frac{\sum x_{1,i}^2 \sum x_{2,i}^2}{\sum x_{1,i}^2} - \frac{\left(\sum x_{1,i}x_{2,i}\right)^2}{\sum x_{1,i}^2}\right]}\sigma^2 \\
&= \frac{1}{\sum x_{2,i}^2 \left[\frac{\sum x_{1,i}^2}{\sum x_{1,i}^2} - \frac{\left(\sum x_{1,i}x_{2,i}\right)^2}{\sum x_{2,i}^2 \sum x_{1,i}^2}\right]}\sigma^2 \\
&= \frac{1}{\sum x_{2,i}^2 \left[1 - \ r_{12}^2\right]}\sigma^2 \\
&= \frac{1}{(1-r_{12}^2)}\frac{1}{\sum x_{2,i}^2}\sigma^2 \qquad (340)
\end{aligned}$$

Variance Inflation Factor in Inexact Multicollinearity
Let correlations between $X_{1,i}$ and $X_{2,i}$ be given by r_{12}. Then Variance inflation factor is $\frac{1}{(1-r_{12}^2)}$

$$\begin{aligned}
var\left(\widehat{\beta}_1\right) &= \frac{\sum x_{2,i}^2}{\sum x_{1,i}^2 \sum x_{2,i}^2 - \left(\sum x_{1,i}x_{2,i}\right)^2}\sigma^2 \\
& \frac{1}{\left[\frac{\sum x_{1,i}^2 \sum x_{2,i}^2}{\sum x_{2,i}^2} - \frac{\left(\sum x_{1,i}x_{2,i}\right)^2}{\sum x_{2,i}^2}\right]}\sigma^2 \\
&= \frac{1}{\sum x_{1,i}^2 \left[\frac{\sum x_{2,i}^2}{\sum x_{2,i}^2} - \frac{\left(\sum x_{1,i}x_{2,i}\right)^2}{\sum x_{1,i}^2 \sum x_{2,i}^2}\right]}\sigma^2 \\
&= \frac{1}{\sum x_{1,i}^2 \left[1 - \ r_{12}^2\right]}\sigma^2 \qquad (341)
\end{aligned}$$

Solutions for Multicollinearity Problem

When Variance is high the standard errors are hish and that makes t-statistics very small and insignificant

$$SE\left(\widehat{\beta}_2\right) = \sqrt{var\left(\widehat{\beta}_2\right)}\ ;\ SE\left(\widehat{\beta}_1\right) = \sqrt{var\left(\widehat{\beta}_1\right)}$$
$$t_{\widehat{\beta}_1} = \frac{\widehat{\beta}_1 - \beta_1}{SE\left(\widehat{\beta}_1\right)};\ t_{\widehat{\beta}_2} = \frac{\widehat{\beta}_2 - \beta_2}{SE\left(\widehat{\beta}_2\right)} \tag{342}$$

.since $0 < r_{12} < 1$ it raises the variance and hence stancard errors and lowers t-values.

1. First detect the pairwise correlations between explalantory variables such $X_{1,i}$ and $X_{3,i}$ be given by r_{12}.
2. Drop highly correlated variables.
3. Adopts Klein's rule of thumb:
4. Compare R_y^2 from overall regression to R_x^2 from auxiliary regression. Determine multicollinearity if $R_x^2 > R_y^2$. Drop highly correlated variables.

1.7.2 Exercise 5

1. Data on income (y), performance indicator (x_1) and quality of workers (x_2) in a certatin reputable company is given as following.

Table 13: Data on income, performace and quality of work

y	3	5	7	6	9	6	7
x_1	1	2	3	4	5	6	7
x_2	5	10	15	20	25	30	35

Fit a regression model $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$ for this company. If any problem suggest remedial measures.

2. Some international macroeconomists argue that the devaluation has expansionary effect on output through its positive impact on exports and negative impacts on imports. Others think that devaluation has contractionary impact on output. As an econometrician you would like to test which one of these two claims bear close relation to the empirical facts. Based on literature review you come up with the following model

$$g_{y,t} = \beta_0 + \beta_1 time + \beta_2\left(\frac{G}{Y}\right) + \beta_3\left[\Delta\ln M - \Delta lmM^*\right] + \beta_4 TOT + \beta_5 RE_t + \varepsilon_t \tag{343}$$

Where $g_{y,t}$ is the growth rate of real output, $time$ is time trend, $\frac{G}{Y}$ is the ratio of government expenditure to GNP, M is the money supply, M^* is expected money supply , TOT is the term of trade as provided by the ratio of indices of price of exports to the prices of imports, RE is the real exchange rate. Terms β_0,β_1,β_2,β_3,and β_4 are unknown coefficients to be estimated. As before is the error term, it has a zero mean and constant variance, $\varepsilon_i \sim N(0, \sigma^2)$.

Relevant data are proveded in juk.xlsx file. Estimate the above parameters and answer following questions studying the regression results.

1. (a) Explain significance of coefficients β_0,β_1,β_2,β_3,$\beta_4$and β_5 in the above model and state whether the estimates are consistent with the economic theory. Is a devaluation, an increase in RE, contractionary or expansionary from the results of this model?

   (b) Explain how you can test three of the following restrictions (1) separately and (2) jointly in this model.

      i. Restriction 1: $\beta_5 = 0$
      ii. Restriction 2: $\beta_2 = 0$ and $\beta_4 = 0$
      iii. Restriction 3: $\beta_3 + \beta_4 = 0$
      iv. Discuss your test statistic for (i) to (iv).

   (c) If the data series used in this model is non-stationary, mention how does it affect the estimates of the parameters? What would you do to correct it?

- Koutsoyiannis A. (1984) Goals of Oligopolistic Firms: An Empirical Test of Competing Hypotheses. Southern Economic Journal, 51, 2, 540-567

1.8 Heteroreskedastity

Consider a linear regression

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad i = 1 \ldots N \tag{344}$$

and assumptions

$$E(\varepsilon_i) = 0 \tag{345}$$

$$E(\varepsilon_i x_i) = 0 \tag{346}$$

$$var(\varepsilon_i) = \sigma^2 \quad for \ \forall \ i \tag{347}$$

$$covar(\varepsilon_i \varepsilon_j) = 0 \tag{348}$$

Then the OLS Regression coefficients are:

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}; \quad \widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \tag{349}$$

Heteroskedasticity occurs when variances of errors are not constant, $var(\varepsilon_i) \neq \sigma_i^2$ variance of errors vary for each i. This is mainly a cross section problem. Main reason for this are

- Learning reduces errors;
  - driving practice, driving errors and accidents
  - typing practice and typing errors,

– defects in productions and improved machines

- Improved data collection: better formulas and goods software
- More heteroscedasticity exists in cross section than in time series data.

Nature of Heteroskedasticity

$$E(\varepsilon_i)^2 = \sigma_i^2 \tag{350}$$

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \tag{351}$$

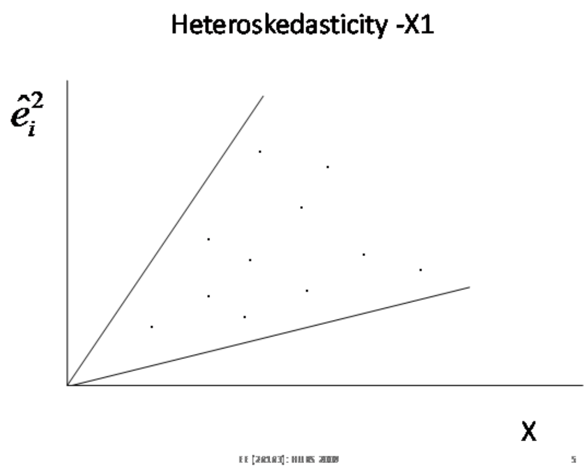
$$E\left(\widehat{\beta}_2\right) = \sum w_i y_i \tag{352}$$

where

$$w_i = \frac{x_i}{\sum x_i^2} = \frac{\left(X_i - \overline{X}\right)}{\sum \left(X_i - \overline{X}\right)^2} \tag{353}$$

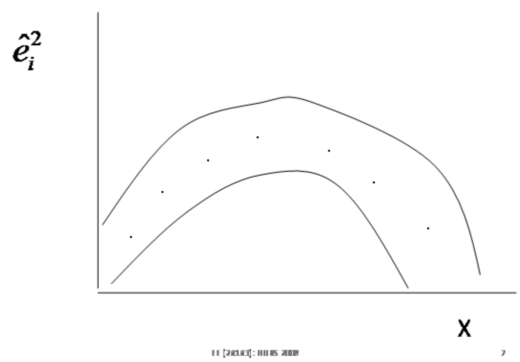
$$Var\left(\widehat{\beta}_2\right) = var\left[\frac{\sum \left(X_i - \overline{X}\right)}{\sum \left(X_i - \overline{X}\right)^2}\right] var(y_i) = \frac{\sum x_i^2 \sigma_i^2}{\left[\sum x_i^2\right]^2} \tag{354}$$

1.8.1 Graphical detection of the heteroskedasticity

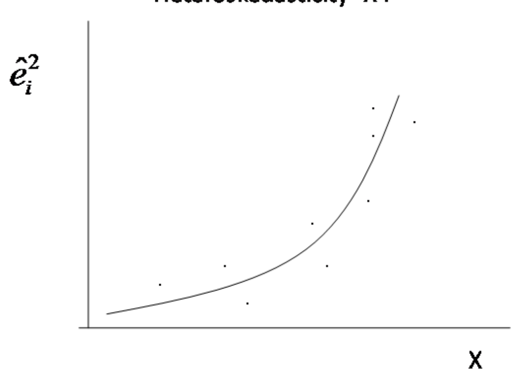


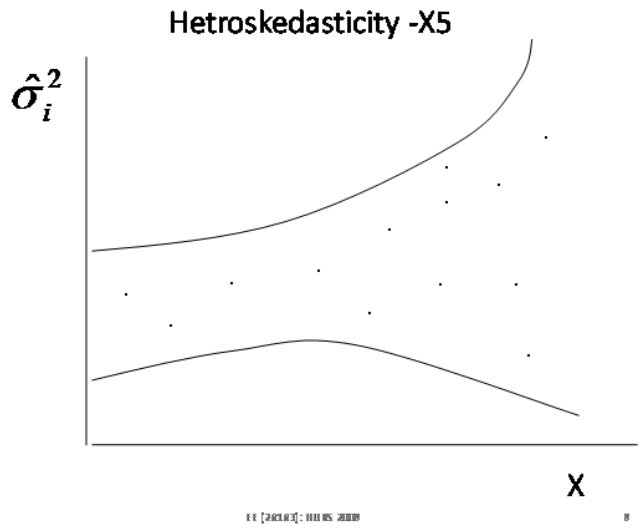
Hetroskedasticity -X2

Hetroskedasticity -X3



Heteroskedasticity -X4





OLS Estimator is still unbiased

$$\widehat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \sum w_i y_i \tag{355}$$

$$E\left(\widehat{\beta}_2\right) = E\left(\sum w_i y_i\right) = E\sum w_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right) \tag{356}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_i\right) + \beta_2 E\left(\sum w_i x_i\right) + E\left(\sum w_i \varepsilon_i\right) \tag{357}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 \tag{358}$$

OLS Parameter is inefficient with Heteroskedasticity

$$E\left(\widehat{\beta}_2\right) = \sum w_i y_i \tag{359}$$

$$E\left(\widehat{\beta}_2\right) = E\left(\sum w_i y_i\right) = E\sum w_i\left(\beta_1 + \beta_2 X_i + \varepsilon_i\right) \tag{360}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_i\right) + \beta_2 E\left(\sum w_i x_i\right) + E\left(\sum w_i \varepsilon_i\right) \tag{361}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 + E\left(\sum w_i \varepsilon_i\right) \tag{362}$$

$$Var\left(\widehat{\beta}_2\right) = E\left[E\left(\widehat{\beta}_2\right) - \beta_2\right]^2 = E\left(\sum w_i \varepsilon_i\right)^2 \tag{363}$$

$$Var\left(\widehat{\beta}_2\right) = E\left(\sum\sum w_i^2 \varepsilon_i^2\right) + \sum\sum cov\left(\varepsilon_i \varepsilon_j\right)^2 \tag{364}$$

$$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left[\sum x_i^2\right]^2} \tag{365}$$

OLS Estimator is inconsistent assymptotically

$$Var\left(\widehat{\beta}_2\right) = \frac{\sum x_i^2 \sigma_i^2}{\left[\sum x_i^2\right]^2} \tag{366}$$

$$\underset{\lim \quad N \to \infty}{Var\left(\widehat{\beta}_2\right)} = \underset{\lim \quad N \to \infty}{\frac{\sum x_i^2 \sigma_i^2}{\left[\sum x_i^2\right]^2}} \Rightarrow \infty \tag{367}$$

1.8.2 Various tests of heteroskedasticity

- Spearman Rank Test
- Park Test
- Goldfeld-Quandt Test
- Glesjer Test
- Breusch-Pagan,Godfrey test
- White Test
- ARCH test

  (See food_hetro.xls excel spreadsheet for some exmaples on how to compute these. Gujarati (2003) Basic Econometrics, McGraw Hill is a good text for Heteroskedasticity; x-hetro test in PcGive).

GLS Solution of the Heteroskedasticity Problem When Variance is Known

$$\frac{Y_i}{\sigma_i} = \frac{\beta_1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i} \quad i = 1 \; ...N \tag{368}$$

Variance with this tranformation equals 1. $var\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{\sigma_i^2}{\sigma_i^2} = 1$
if

$$\sigma_i^2 = \sigma^2 X_i \tag{369}$$

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{\varepsilon_i}{X_i} \; ; \quad var\left(\frac{\varepsilon_i}{x_i}\right) = \frac{\sigma^2 x_i^2}{x_i^2} = \sigma^2 \tag{370}$$

In matrix notation

$$\beta_{OLS} = \left(X'X\right)^{-1}\left(X'Y\right) \tag{371}$$

$$\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right) \tag{372}$$

Ω^{-1} is inverse of variance covariance matrix.

Spearman rank test of heteroskedactity

$$r_s = 1 - 6 \times \frac{\sum\limits_i d_i^2}{n\left(n^2-1\right)} \tag{373}$$

- steps:
- run OLS of y on x.
- obtain errors e
- rank e and y or x
- find the difference of the rank
- use t-statistics if ranks are significantly different assuming $n > 8$ and rank correlation coefficient $\rho = 0$.

$$t = 1 - 6 \times \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \quad with \quad df \quad (n-2) \tag{374}$$

- If $t_{cal} > t_{crit}$ there is heteroskedasticity.

Glesjer Test of heteroskedasticity

- Model
$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ldots N \tag{375}$$
- There are a number of versions of it:

$$|e_i| = \beta_1 + \beta_2 X_i + v_i \tag{376}$$

$$|e_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i \tag{377}$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i \tag{378}$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i \tag{379}$$

$$|e_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i \tag{380}$$

$$|e_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i \tag{381}$$

- In each case test H0: $\beta_i = 0$ against HA: $\beta_i \neq 0$. If is significant then that is the evidence of heteroskedasticity.

White test
White test of heteroskedasticity is more general test

1. This is a more general test

2. Model $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i$
3. Run an initial OLS to this and get $\widehat{e}_i^2$. Then regress $\widehat{e}_i^2$ to squares and cross products of $X_{j,i}$
4. $\widehat{e}_i^2 = \alpha_0 + \alpha_1 X_{1,i} + \alpha_2 X_{2,i} + \alpha_3 X_{3,i} + \alpha_4 X_{1,i}^2 + \alpha_5 X_{2,i}^2 + \alpha_6 X_{3,i}^2 + \alpha_7 X_{1,i} X_{2,i} + \alpha_8 X_{2,i} X_{3,i}. + \varepsilon_i$
5. Compute the test statistics; H0: no heteroskedasticity.
6. $n.R^2 \sim \chi^2_{df}$
7. Again if the calculated χ^2_{df} is greater than table value there is an evidence of heteroskedasticity.

Park test of heteroskedasticity

- Model
$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ ...N \tag{382}$$
- Error square:
$$\sigma_i^2 = \sigma^2 X_i^\beta e_i^{v_i} \tag{383}$$
- Or taking log
$$ln\sigma_i^2 = \ln \sigma^2 + \beta_2 X_i + v_i \tag{384}$$
- steps : run the OLS regression for (Y_i) and get the estimates of error terms $(e_i\ )$.
- Square e_i , and then run a regression of lne_i^2 with x variable. Do t-test H0: $\beta_2 = 0$ against HA: $\beta_2 \neq 0$. If is significant then that is the evidence of heteroskedasticity.

Goldfeld-Quandt test of heteroskedasticity

- Model
$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ ...N \tag{385}$$
- Steps:
  - Rank observations in ascending order of one of the x variable
  - Omit c numbers of central observations leaving two groups $\frac{N-C}{2}$ with number of osbervations
  - Fit OLS to the first $\frac{N-C}{2}$ and the last $\frac{N-C}{2}$ observations and find sum of the squared errors from both of them.
  - Set hypothesis $\sigma_1^2 = \sigma_2^2$ against $\sigma_1^2 \neq \sigma_2^2$.
  - compute $\lambda = \frac{ERSS_2/df2}{ERSS_1/df1}$.
  - It follows F distribution.

Breusch-Pagan,Godfrey test of heteroskedasticity
$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + ... + \beta_k X_{k,i} + \varepsilon_i \quad i = 1 \ ...N$

- run OLS and obtain error squares

- Obtain average error square $\widehat{\sigma}^2 = \frac{\sum_i e_i^2}{n}$ and $p_i = \frac{e_i^2}{\widehat{\sigma}^2}$
- regress p_i on a set of explanatory variables
- $p_i = \alpha_0 + \alpha_1 X_{1,i} + \alpha_2 X_{2,i} + \alpha_3 X_{3,i} + ... + \alpha_k X_{k,i} + \varepsilon_i$
- obtain squares of explained sum (EXSS)
- $\theta = \frac{1}{2}(EXSS)$
- $\theta = \frac{1}{m-1}(EXSS) \sim \chi^2_{m-1}$
- $H_0 : \alpha_0 = \alpha_1 = \alpha_2 = \alpha_3 = .. = \alpha_k = 0$
- No heteroskedasticity and $\sigma_i^2 = \alpha_1$ a constant. If calculated χ^2_{m-1} is greater than table value there is an evidence of heteroskedasticity.

ARCH test of heteroskedasticity

Engle (1987) autoregressive conditional heteroskedasticy (ARCH): more useful for time series data

Model $Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + ... + \beta_k X_{k,t} + e_t$

1. - $\varepsilon_t \sim N\left(0, \left(\alpha_0 + \alpha_2 e_{t-1}^2\right)\right)$

   $$\sigma_t^2 = \alpha_0 + \alpha_2 e_{t-1}^2 \tag{386}$$

   - Here σ_t^2 not observed. Simple way is to run OLS of Y_t and get $\widehat{e}_t^2$
   - ARCH (1)
   - $\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + v_t$
   - ARCH (p)
   - $\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{1-1}^2 + \alpha_4 \widehat{e}_{1-1}^2 + .. + \alpha_p \widehat{e}_{1-p}^2 + v_t$
   - Compute the test statistics
   - $n.R^2 \sim \chi^2_{df}$
2. Again if the calculated χ^2_{df} is greater than table value there is an evidence of ARCH effect and heteroskedasticity.
3. Both ARCH and GARCH models are estimated using iterative Maximum Likelihood procedure.

GARCH tests of heteroskedasticity

Bollerslev's generalised autoregressive conditional heteroskedasticy (GARCH) process is more general

- GARCH (1)

  $$\sigma_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \beta \sigma_{t-1}^2 + v_t \tag{387}$$

- GARCH (p,q)
- $\sigma_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{t-2}^2 + \alpha_4 \widehat{e}_{t-3}^2 + .. + \alpha_p \widehat{e}_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + ..\beta_q \sigma_{t-q}^2 + .. + v_t$

- Compute the test statistics $n.R^2 \sim \chi^2_{df}$
- Sometimes written as
- $h_t = \alpha_0 + \alpha_2 \widehat{e}^2_{t-1} + \alpha_3 \widehat{e}^2_{t-2} + \alpha_4 \widehat{e}^2_{t-3} + .. + \alpha_p \widehat{e}^2_{t-p} + \beta_1 h_{t-1} + \beta_2 h_{t-2} + ..\beta_q h_{t-q} + .. + v_t$
- $where \quad h_t = \sigma^2_t$
- Various functional forms of h_t
- $h_t = \alpha_0 + \alpha_2 \widehat{e}^2_{t-1} + \beta_1 \sqrt{h_{t-1}} + v_i$ or $h_t = \alpha_0 + \alpha_2 \widehat{e}^2_{t-1} + \sqrt{\beta_1 h_{t-1} + \beta_2 h_{t-2}} + v_i$
- Both ARCH and GARCH models are estimated using iterative Maximum Likelihood procedure. Volatility package in PcGive estimates ARCH-GARCH models; Eviews, STATA or RATS also have these routines.

1.8.3 Exercise 6

Take a simple linear regression model of the following form.

$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ ...N \tag{388}$$

Where the variance of the error term differs for different observations of X_i.

1. (a) Discuss how the graphical method be used to detect the heteroskedasticity.
   (b) Prove that parameters $\beta_1$and β_2 are still unbiased.
   (c) Analyse consequences of heteroskedasticity on the BLUE properties of the OLS estimators.
   (d) Discuss how the Goldfeld and Quandt and Glesjer tests can be used to determine existence of the heteroskedasticity problem.
   (e) Illustrate procedure for the White test of heteroskedasticity.
   (f) Illustrate any two remedial measures of removing the heteroskedasticity when the variance and is known and when it is unknown.
   (g) From a sample of 6772 observations on pay work-hours and taxes contained in PAYHRTX.XLS determine whether heteroskedasticity exists or not on the basis of cross section estimates from the the PcGive. Feel free to use Shazam if you know and prefer it.
   (h) From a sample of 201 counties of Great Britain contained in Unempl_pay_counties.csv regress work-hours on annual pay and determine whether heteroskedasticity is present in this estimation using the White test.
   (i) Suggest remedial measures to remove heteroskedasticity in models like above.
   (j) Explains concepts of ARCH and GARCH models briefly.

Do more exercises with cross section data such as the annual population survey, customer satisfacton survey, job satisfaction survey, Family Expenditure Survey or other surveys you know.

1.9 Autocorrelation

Consider a linear regression

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \quad t = 1 \ldots T \tag{389}$$

Classical assumptions

$$E(\varepsilon_t) \quad = 0 \tag{390}$$

$$E(\varepsilon_t x_t) \quad = 0 \tag{391}$$

$$var(\varepsilon_t) = \sigma^2 \quad for \; \forall \; t \;\; covar(\varepsilon_t \varepsilon_{t-1}) = 0 \tag{392}$$

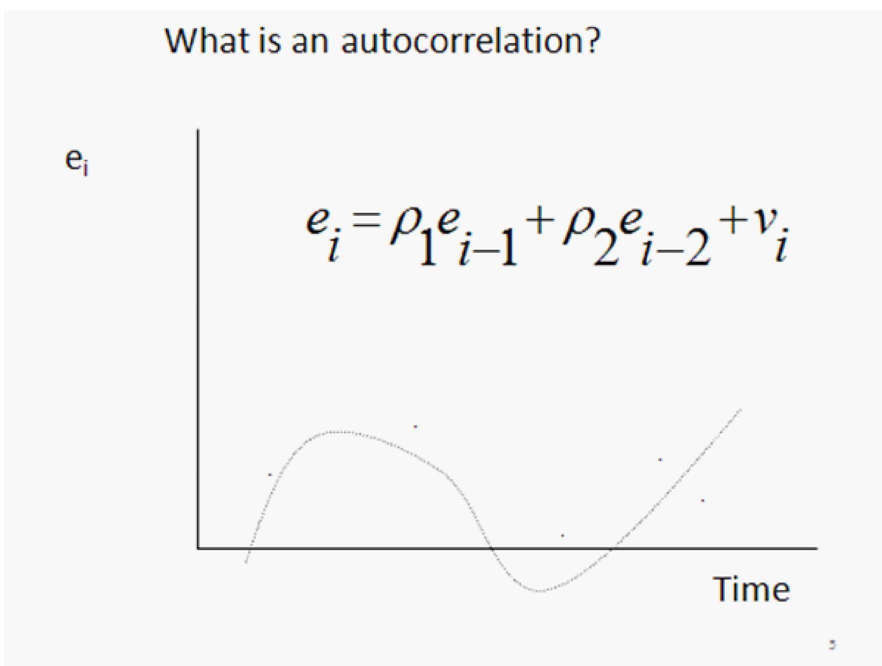
In presence of autocorrelation (first order)

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \tag{393}$$

Then the OLS Regression coefficients are:

$$\widehat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2}; \quad \widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \quad ; \widehat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2} \tag{394}$$

Autocorrelation occurs when covariances of errors are not zero, $covar(\varepsilon_t \varepsilon_{t-1}) \neq 0$ covariance of errors are nonnegative This is mainly a problem observed in time series data.



Causes of autocorrelation

- inertia , specification bias, cobweb phenomena
- manipulation of data

Consequences of autocorrelation

1. (a) Estimators are still linear and unbiased, but
   (b) they there not the best, they are inefficient.

Remedial measures

1. (a) When ρ is known - transform the model
   (b) When ρ is unknown estimate it and transform the model

1.9.1 Nature of autocorrelation

$$\widehat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \tag{395}$$

$$E\left(\widehat{\beta}_2\right) = \sum w_t y_t \tag{396}$$

where

$$E\left(\varepsilon_t\right)^2 = \sigma^2 \tag{397}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 + E\left(\sum w_t \varepsilon_t\right) \tag{398}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_t\right) + \beta_2 E\left(\sum w_t x_t\right) + E\left(\sum w_t \varepsilon_t\right) \tag{399}$$

$$Var\left(\widehat{\beta}_2\right) = E\left[E\left(\widehat{\beta}_2\right) - \beta_2\right]^2 = E\left(\sum w_t \varepsilon_t\right)^2 \tag{400}$$

$$Var\left(\widehat{\beta}_2\right) = \frac{1}{\sum x_t^2}\sigma^2 + \sum\sum cov\left(\varepsilon_t \varepsilon_{t-1}\right) \tag{401}$$

OLS Estimator is still unbiased

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \tag{402}$$

$$\widehat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} = \sum w_t y_t \tag{403}$$

$$E\left(\widehat{\beta}_2\right) = E\left(\sum w_t y_t\right) = E\sum w_t\left(\beta_1 + \beta_2 X_t + \varepsilon_t\right) \tag{404}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_t\right) + \beta_2 E\left(\sum w_t x_t\right) + E\left(\sum w_t \varepsilon_t\right) \tag{405}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 \tag{406}$$

1.9.2 OLS Parameters are inefficient with autocorrelation

$$E\left(\widehat{\beta}_2\right) = \sum w_t y_t \tag{407}$$

$$E\left(\widehat{\beta}_2\right) = E\left(\sum w_t y_t\right) = E\sum w_t\left(\beta_1 + \beta_2 X_t + \varepsilon_t\right) \tag{408}$$

$$E\left(\widehat{\beta}_2\right) = \beta_1 E\left(\sum w_t\right) + \beta_2 E\left(\sum w_t x_t\right) + E\left(\sum w_t \varepsilon_t\right) \tag{409}$$

$$E\left(\widehat{\beta}_2\right) = \beta_2 + E\left(\sum w_t \varepsilon_t\right) \tag{410}$$

$$Var\left(\widehat{\beta}_2\right) = E\left[E\left(\widehat{\beta}_2\right) - \beta_2\right]^2 = E\left(\sum w_t \varepsilon_t\right)^2 \tag{411}$$

$$Var\left(\widehat{\beta}_2\right) = E\left(\sum\sum w_t^2 \varepsilon_t^2\right) + 2\sum\sum w_t w_{t-1} cov\left(\varepsilon_t \varepsilon_{t-1}\right) \tag{412}$$

$$Var\left(\widehat{\beta}_2\right) = \frac{1}{\sum x_t^2}\sigma^2\left[1 + 2\frac{\sum x_t x_{t-1}}{\left[\sum x_t^2\right]}\frac{cov\left(\varepsilon_t \varepsilon_{t-1}\right)}{\sqrt{var\left(\varepsilon_t\right)}}\right] \because var\left(\varepsilon_t\right) = var\left(\varepsilon_{t-1}\right) \tag{413}$$

$$Var\left(\widehat{\beta}_2\right) = \frac{1}{\sum x_t^2}\sigma^2\left[\begin{array}{c} 1 + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^1 + \\ +2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^2 + .. + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^s \end{array}\right] \tag{414}$$

OLS Estimator is inconsistent assymptotically

$$Var\left(\widehat{\beta}_2\right) = \frac{1}{\sum x_t^2}\sigma^2\left[\begin{array}{c} 1 + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^1 + \\ +2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^2 + .. + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^s \end{array}\right] \tag{415}$$

$$\underset{\lim \quad N \to \infty}{Var\left(\widehat{\beta}_2\right)} = \frac{1}{\sum x_t^2}\sigma^2\left[\begin{array}{c} 1 + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^1 + \\ +2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^2 + .. + 2\frac{\sum(x_t-\overline{x})(x_{t-1}-\overline{x})}{\sum x_t^2}\rho^s \end{array}\right] \Rightarrow \infty \tag{416}$$

1.9.3 Durbin-Watson test of autocorrelation

$$d = \frac{\sum\limits_{t=1}^{T}\left(e_t - e_{t-1}\right)^2}{\sum\limits_{t=1}^{T} e_t^2} \tag{417}$$

$$d = \frac{\sum\limits_{t=1}^{T}\left(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2\right)}{\sum\limits_{t=1}^{T} e_t^2} = 2\left(1-\rho\right); \quad \because \sum_{t=1}^{T} e_t^2 \simeq \sum_{t=2}^{T} e_{t-1}^2 \tag{418}$$

Autocorrelation coefficient is given by:

$$\rho = \frac{\sum_{t=1}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} \tag{419}$$

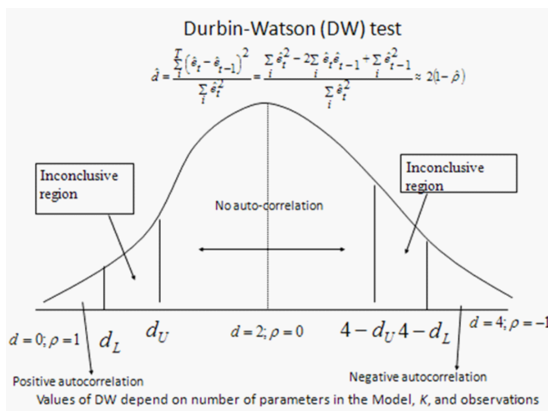
Autocorrelation and Durbin-Watson Statistics

$$d = 2\left(1 - \rho\right) \tag{420}$$

$$\rho = 0 \Longrightarrow d = 2 \tag{421}$$

$$\rho = -1 \Longrightarrow d = 4 \tag{422}$$

Durbin-Watson Distribution



Transformation of the model in the presence of autocorrelation
when autocorrelation coefficient is known

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \quad t = 1 \ldots T \tag{423}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \tag{424}$$

$$Y_t - \rho Y_{t-1} = \left(\beta_1 - \rho\beta_1\right) + \beta_2\left(X_t - \rho X_{t-1}\right) + \varepsilon_t - \rho\varepsilon_{t-1} \tag{425}$$

$$Y_t^* = \beta_1^* + \beta_2 X_t^* + \varepsilon_t^* \tag{426}$$

Apply OLS in this transformed model β_1^* and β_2 will have BLUE properties.

When autocorrelation coefficient is unknown, this method is similar to the above ones, except that it involves multiple iteration for estimating ρ . Steps are as following:

1. Get estimates $\widehat{\beta}_1$ and $\widehat{\beta}_2$ from the original model; get error terms $\widehat{e}_i$ and estimate $\widehat{\rho}$
2. Transform the original model multiplying it by $\widehat{\rho}$ and by taking the first difference,
3. Estimate $\widehat{\widehat{\beta}}_1$ and $\widehat{\widehat{\beta}}_2$ from the transformed model and get errors $\widehat{\widehat{e}}_i$ of this transformed model
4. Then again estimate $\widehat{\widehat{\rho}}$ and use those values to transform the original model as

$$Y_t - \widehat{\rho} Y_{t-1} = (\beta_1 - \widehat{\rho}\beta_1) + \beta_2 (X_t - \widehat{\rho} X_{t-1}) + \varepsilon_t - \widehat{\rho}\varepsilon_{t-1} \tag{427}$$

5. Continue this iteration process until $\widehat{\widehat{\rho}}$ converges.

PcGive suggests using differences in variables. Diagnos /ACF options in OLS in Shazam will generate these iterations.

1.9.4 GLS to solve autocorrelation

In matrix notation

$$\beta_{OLS} = (X'X)^{-1} (X'Y) \tag{428}$$

$$\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} \left(X'\Omega^{-1}Y\right) \tag{429}$$

Ω^{-1} is inverse of variance covariance matrix.

Generalised Least Square

Take a regression

$$Y = X\beta + e \tag{430}$$

Assumption of homoskedasticity and no autocorrelation are violated

$$var(\varepsilon_i) \neq \sigma^2 \quad for \ \forall \ i \tag{431}$$

$$covar(\varepsilon_i \varepsilon_j) \neq 0 \tag{432}$$

The variance covariance of error is given by

$$\Omega = E(ee') = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & .. & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & .. & \sigma_{2n} \\ : & : & : & : \\ \sigma_{n1} & \sigma_{n2} & .. & \sigma_n^2 \end{bmatrix} \tag{433}$$

$$Q'\Omega Q = \Lambda \tag{434}$$

Generalised Least Square

$$\Omega = Q\Lambda Q' = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q' \tag{435}$$

$$P = Q\Lambda^{\frac{1}{2}} \tag{436}$$

$$P'\Omega P = I \quad ; \quad P'P = \Omega^{-1} \tag{437}$$

Transform the model

$$PY = \beta PX + Pe \tag{438}$$

$$Y^* = \beta X^* + e^* \tag{439}$$

$Y^* = PY \quad X^* = PX \quad and \quad e^* = Pe$ $\qquad \beta_{GLS} = (X'P'PX)^{-1}(X'P'PY)$

$$\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right) \tag{440}$$

1.9.5 Exercise 7

Suppose that you are estimating a log linear consumption function of the following form:

$$\ln(C_t) = \beta_0 + \beta_1 \ln(Y_t) + \beta_2 \ln(P_t) + \varepsilon_t \quad t = 1 \ldots T \tag{441}$$

where C, Y and P are consumption, income and prices and ε_t is the random error term. Use information in conyp.xls to estimate unknown parametersβ_0,$\beta_1$and β_2 and answer following questions using these results.

1. (a) What are the estimates of $\beta_1$and β_2? Do these estimates have signs as you expected and why?
   (b) Does the Durbin-Watson Statistic show evidence of autocorrelation in the model? If so how does it affect the properties of the OLS estimators of $\beta_1$and β_2?
   (c) What is the 95 and 90 percent of confidence interval estimate of $\beta_1$and β_2?
   (d) How well does this model can explain variation in consumption? How do you decide overall fit of this model? What statistics do you use to decide at least there is one significant variable in the model?

2. Consider a simple linear regression model.

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad t = 1 \ldots T \tag{442}$$

Now assume that errors are correlated to each other over time with AR(1) process as:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \tag{443}$$

where v_t is identically and normally distributed error term with zero mean and constant variance, $v_t \sim N(0, \sigma^2)$.

1. (a) Illustrate how the graphical method can be applied to detect autocorrelation in a simple regression model like above?
   (b) What are consequences of autocorrelation in a regression model? Show how the existence of such autocorrelation among the error terms affects the BLUE properties of the OLS estimators.
   (c) Define and derive the Durbin-Watson test statistics. Show how it can test for existence or non existence of autocorrelation in a given estimation?
   (d) How the autocorrelation can be removed if the ρ is known?

(e) What is a spurious regression? Why does it arise and how does it affect the usefulness of estimation from an OLS regression? What can be done to correct it?

Application:
Read data on growth rate of per capita GDP, exchange rate and inflation rates from the www.imf.org for year 1980 to 2003 for China, India, South Africa, UK, USA and Brazil as contained in PERCAP6.GLS. Test whether inflation and the exchange rate are the significant variables in explaining the growth rate of per capita output (in PPP) in these economies. Determine whether heteroskedasticity and autocorrelation exist in this regression using PcGive. Feel free to use Shazam if you know and prefer it. Suggest a remedy for autocorrelation in a model like this.

1.10 Time Series

Time series models aim to explain the data generating process for $\{y_t\}_{-\infty}^{\infty} = \left\{ \begin{array}{c} y_{-\infty}.....y_{-1}.y_0.y_1.y_2... \\ .....y_T.y_{T+1}.y_{T+1}..... \end{array} \right\}$

A Time series consists of trend, cycle, season and irregular component

$$Y = T \times C \times S \times I \tag{444}$$

In a simple method the moving average gives $T \times C$ components and is used to isolate the $S \times I$ components. For instance for a 12 monthly moving average

$$\overline{Y}_i = \frac{1}{12}\left(Y_1 + Y_2 + ... + Y_{12}\right) \tag{445}$$

$$S \times I = \frac{T \times C \times S \times I}{T \times C} = \frac{Y_i}{\overline{Y}_i} = z_t \tag{446}$$

Now to isolate the Irregular component I from $S \times I$ take out the seasonal elements from z_t assuming monthly data for 5 years (60 observations) compute the seasonal indices as following:

$$Month1 : \overline{z}_1 = \frac{1}{5}\left(z_1 + z_{13} + z_{25} + z_{39} + z_{48}\right) \tag{447}$$

$$Month2 : \overline{z}_2 = \frac{1}{5}\left(z_2 + z_{14} + z_{26} + z_{40} + z_{49}\right) \tag{448}$$

$$Month3 : \overline{z}_3 = \frac{1}{5}\left(z_3 + z_{15} + z_{26} + z_{41} + z_{50}\right) \tag{449}$$

....

$$Month11 : \overline{z}_{11} = \frac{1}{5}\left(z_{11} + z_{23} + z_{35} + z_{47} + z_{59}\right) \tag{450}$$

$$Month12 : \overline{z}_{12} = \frac{1}{5}\left(z_{12} + z_{24} + z_{36} + z_{46} + z_{60}\right) \tag{451}$$

Deseasonalisation of data $Y_i^d = \frac{Y_i}{\overline{z}_i}$ and irregular component should be $i = \frac{z_t}{\overline{z}_i}$.

Trends:

Simple extrapolation

$$Y_t = c_1 + c_2 t \tag{452}$$

Exponential growth

$$Y_t = Ae^{rt} \tag{453}$$

Autoregressive model

$$Y_t = c_1 + c_2 Y_{t-1} \tag{454}$$

Log trend

$$\ln(Y_t) = c_1 + c_2 \ln(Y_{t-1}) \tag{455}$$

Quadratic trends:

$$Y_t = c_1 + c_2 t + c_3 t^2 \tag{456}$$

Logistic trend:

$$Y_t = \frac{1}{k + bt} \quad b > 1 \tag{457}$$

$$Y_t = e^{k_1 - \frac{k_2}{t}} \tag{458}$$

$$ln(Y_t) = k_1 - \frac{k_2}{t} \tag{459}$$

auto lagged with declining weights $\alpha < 1$

$$Y_t = \alpha Y_{t-1} + \alpha(1-\alpha) Y_{t-2} + \alpha(1-\alpha)^2 Y_{t-2} + ... + \alpha(1-\alpha)^n Y_{t-2} \tag{460}$$

Forecasting forward with these models is obvious.

1.10.1 Time series process

Simplest of these is a trend model

$$Y_t = \beta t + \varepsilon_t \tag{461}$$

with mean $E(Y_t) = \beta t$ and variance $E(Y_t - \beta t)^2 = E(\varepsilon_t)^2 = \sigma_\varepsilon^2$
Or it could have been just a constant plus a Gaussian white noise $\varepsilon_t \sim N(0, \sigma^2)$ as:

$$Y_t = \mu + \varepsilon_t \tag{462}$$

with mean $E(Y_t) = \mu$ and variance $E(Y_t - \mu)^2 = E(\varepsilon_t)^2 = \sigma_\varepsilon^2$
Autocovariance of $\{y_t\}_{-\infty}^{\infty}$ for I realisations is

$$\gamma_{tj} = E(Y_t - \mu) E(Y_{t-j} - \mu) = E(\varepsilon_t) E(\varepsilon_{t-j}) = 0 \quad for\ j \neq 0 \tag{463}$$

Stationarity

when neither mean μ nor the autocovariance γ_{ij} depend on time t then the Y_t is covariance stationary or weakly stationary.

$$E(Y_t) = \mu \; for \; \forall \; t \tag{464}$$

$$E\left(Y_t - \mu\right) E\left(Y_{t-j} - \mu\right) = \gamma_j \quad for \text{ any } t \;\; and \; j = \{^{\sigma_\varepsilon^2 \;\; for \; j=0}_{0 \; for \; j \neq 0} \tag{465}$$

For instance 836 is stationary while 835 not covariance stationary because its mean βt is function of time.

If the process is stationary γ_j is the same for any value of t $\gamma_j = \gamma_{-j}$

$$\gamma_j = E\left(Y_{t+j} - \mu\right) E\left(Y_{(t+j)-j} - \mu\right) = E\left(Y_{t+j} - \mu\right) E\left(Y_t - \mu\right) = E\left(Y_t - \mu\right) E\left(Y_{t+j} - \mu\right) \; = \gamma_{-j} \tag{466}$$

1.10.2 Stationarity

What is a stationary variable?

When its mean and variance are constant.

$$E\left(Y_t\right) = \mu \tag{467}$$

$$var\left(Y_t\right) = \sigma^2 \tag{468}$$

When mean and variances are not constant, that variable is non-starionary, for instance a random walk

$$Y_t = Y_{t-1} + \varepsilon_i \quad t = 1 \; ...T \tag{469}$$

In an autoregressive model

$$Y_t = \rho Y_{t-1} + \varepsilon_i \quad t = 1 \; ...T \tag{470}$$

if the autocorrlation coefficient $\rho = 1$ then it becomes a random walk. This variable is non-stationary.

$$Y_t = \sum_{s=1}^{\infty} \rho^s \varepsilon_{t-s} \tag{471}$$

Current realisations are accumulation of past errors.

Prove that variance of this is .

$$var\left(Y_t\right) = t.\sigma^2 \tag{472}$$

Regression among non-stationary variables becomes spurious unless they are cointegrated.

1.10.3 Unit root and order of integration

A Non-Stationary variable can be made stationary by taking first difference as:

$$\Delta Y_t = Y_t - Y_{t-1} \tag{473}$$

If a variable becomes stationary by taking the first difference it is said to be intergrated of order one

$$I\left(1\right) \tag{474}$$

If it becomes stationary after differencing d time then it is called $I\left(d\right)$ variable.

Dickey-Fuller and Phillip-Perron unit root tests are used to determine stationarity of a variable.

$$Y_t = \rho Y_{t-1} + \varepsilon_i \tag{475}$$

1.10.4 Level, drift, trend and lag terms in unit root test

Dickey-Fuller and Phillip-Perron unit root tests are used to determine stationarity of a variable.

$$Y_t = \rho Y_{t-1} + \varepsilon_i \tag{476}$$

$$\Delta Y_t = \left(\rho - 1\right) Y_{t-1} + \varepsilon_i; \qquad \Delta Y_t = \gamma Y_{t-1} + \varepsilon_i; \tag{477}$$

Random walk with drift

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \varepsilon_i \tag{478}$$

trend stationary

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \varepsilon_i \tag{479}$$

Augmented Dickey-Fuller test

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \sum_{i=1}^{m} \rho^s \Delta Y_{t-i} + \varepsilon_i \tag{480}$$

Cointegration in a regression

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \tag{481}$$

First do the regression and then estimate the error as

$$\widehat{\varepsilon}_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t \tag{482}$$

Y_t and X_t are cointegrated if the estimated error is stationary $\widehat{\varepsilon}_t \sim I\left(0\right)$

$$\widehat{\varepsilon}_t = \rho \widehat{\varepsilon}_{t-1} + \varepsilon_t \tag{483}$$

if $\rho < 1$ the error $\widehat{\varepsilon}_t$ is stationary and Y_t and X_t are cointegrated. They have a long run relationship.

When variables are cointegrated there is an error correction mechanism.

$$Y_t = \varphi_2 X_t + \epsilon_t \tag{484}$$

$$Y_t = X_t + \epsilon_t \ ; \qquad \varphi_2 = 1 \tag{485}$$

Cointegration: Engle-Granger Representation Theorem

$$\epsilon_t \ = Y_t - X_t \tag{486}$$

For test of cointegration

$$\Delta\epsilon_t \ = \gamma\epsilon_{t-1} \ + u_t \tag{487}$$

$$\Delta\left(Y_t - X_t\right) = \gamma\left(Y_{t-1} - X_{t-1}\right) + u_t \tag{488}$$

$$\Delta Y_t = \Delta X_t + \gamma\left(Y_{t-1} - X_{t-1}\right) + u_t \tag{489}$$

This is an error correction model.Term $\gamma\left(Y_{t-1} - X_{t-1}\right)$ gives the adjustment towards the long run equilibrium and ΔX_t denotes the short run impact.

H_0 : No cointegration; t- statistics can be used instead of DF test in error correction model.

Granger Causality Test Estimate the following model where M_t is money Y_t is GDP and test the causality as below:

$$Y_t = \sum_{i=1}^{n}\alpha_i M_{t-i} + \sum_{j=1}^{m}\beta_j Y_{t-j} + u_{1,t} \tag{490}$$

$$M_t = \sum_{i=1}^{n}\lambda_i M_{t-i} + \sum_{j=1}^{m}\delta_j Y_{t-j} + u_{2,t} \tag{491}$$

Unidirection causality from M_t to Y_t requires $\sum_{i=1}^{n}\alpha_i \neq 0$ and $\sum_{j=1}^{m}\delta_j = 0$

Unidirection causality from Y_t to M_t requires $\sum_{j=1}^{m}\delta_j \neq 0$ and $\sum_{i=1}^{n}\alpha_i = 0$

Bilateral causality between Y_t to M_t requires $\sum_{i=1}^{n}\alpha_i \neq 0$ and $\sum_{j=1}^{m}\delta_j \neq 0$

Independence of Y_t to M_t from each other $\sum_{i=1}^{n}\alpha_i = 0$ and $\sum_{j=1}^{m}\delta_j = 0$

1.10.5 Exercise 8

Stationarity, Unit Root and Cointegarion

1. Study the monthly data on unemployment rate and inflation since 1972:1 to 2004:8 as given in "unmnth.xls" file. Use GiveWin PcGive to

- Draw diagrams to represent the rates of unemployment among males and females and the RPI over this period.
- Ascertain whether unit root exists in the overall unemployment rate, URT and RPI at 5% and 1% level of significance in level, in log and in the first difference of these series.
- Detrend the data with Hodrik-Prescott filter and conduct stochastic volatility tests.

2. Regress unemployment rate on inflation rate in levels and in the first differences. Test whether these series are cointegrated using the Engle-Granger procedure. (hint: stationarity of residuals).

3. The time series and represent the underlying data generating processes (DGP) of consumption $\{C_t\}$ and income $\{Y_t\}$. Answer the following questions regarding the properties these series.

   (a) What is meant by saying that$\{C_t\}$ and $\{Y_t\}$ are stationary series? Why is it important that the series are stationary for a robust regression analysis?

   (b) How do you determine whether $\{C_t\}$ and $\{Y_t\}$ are stationary series, or not?

   (c) Analyse the properties of these series when they follow a random walk, or have a unit root.

   (d) What is the meaning of the order of integration in this respect? Discuss any three different methods of checking for stationarity.

   (e) What is the meaning of cointegration between the series and ? How would you decide whether these series$\{C_t\}$ and $\{Y_t\}$ are co-integrated, or not?

   (f) If the original series $\{C_t\}$ and $\{Y_t\}$ are not co-integrated, what transformation can be applied to achieve co-integration? How do you decide the order of co-integration?

   (g) Use time series of consumption and income contained in Quarterly_cons.xls. Determine the order of integration for both consumption and income. Is there an evidence of cointegration between consumption and income in levels or in the first differences?

1.11 Restricted Least Square

Restrictions in Multiple Regression: Restricted Least Square Estimation (Judge-Hill-Griffith-Lutkopohl-Lee (1988): 236)

OLS procedure to minimise the sum of squared error terms.

$$\begin{aligned}
\underset{\beta}{Min}\ S(\beta) &= e'e = (Y-\beta X)'(Y-\beta X) \\
&= Y'Y - Y'(\beta X) - (\beta X)'Y + (\beta X)'(\beta X) \qquad (492) \\
&= Y'Y - 2\beta X'Y + (\beta X)'(\beta X) \qquad (493)
\end{aligned}$$

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2\widehat{\beta}X'X = 0 \Longrightarrow \widehat{\beta} = (X'X)^{-1} X'Y \tag{494}$$

Imposing a restriction involves constrained optimisation with a Lagrange multiplier.

$$\begin{aligned} L &= e'e + 2\lambda\left(r' - \beta' R'\right) = (Y - \beta X)'(Y - \beta X) + 2\lambda\left(r' - \beta' R'\right) \\ &= Y'Y - 2\beta X'Y + (\beta X)'(\beta X) + 2\lambda\left(r' - \beta' R'\right) \end{aligned} \tag{495}$$

Partial derivation of this constrained minimisation function (Lagrangian function) wrt β and λ yields

$$\frac{\partial L}{\partial \beta} = -2X'X + 2X'Xb - 2\lambda R' = 0 \tag{496}$$

$$\frac{\partial L}{\partial \lambda} = -2(r - Rb) = 0 \tag{497}$$

$$X'Xb = X'Y + \lambda R' \tag{498}$$

$$b = (X'X)^{-1} X'Y + (X'X)^{-1} R'\lambda \tag{499}$$

$$b = \widehat{\beta} + (X'X)^{-1} R'\lambda \tag{500}$$

This is the restricted least square estimator but need still to be solved for λ. For that multiply the above equation both sides by R

$$Rb = R\widehat{\beta} + R(X'X)^{-1} R'\lambda \tag{501}$$

$$\lambda = \left[R(X'X)^{-1} R'\right]^{-1} \left[Rb - R\widehat{\beta}\right] \tag{502}$$

$$\lambda = \left[R(X'X)^{-1} R'\right]^{-1} [r - Rb] \tag{503}$$

$$b = \widehat{\beta} + (X'X)^{-1} R'\lambda = \widehat{\beta} + (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1} [r - Rb] \tag{504}$$

Thus the restricted least square estimator is a linear function of the restriction, $[r - Rb]$.

$$E(b) = E\left(\widehat{\beta}\right) + (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1} [r - RE(b)] \tag{505}$$

$$E(b) = E\left(\widehat{\beta}\right) \tag{506}$$

For variance we need to use property of an idempotent matrix AA=A.
Such as

$$A = \begin{bmatrix} 0.4 & 0.8 \\ 0.3 & 0.6 \end{bmatrix} \tag{507}$$

Recall in unrestricted case $\widehat{\beta} = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'e$

$$E(b) - \beta = (X'X)^{-1} X'e + (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1} \left[r - RE(b) - R(X'X)^{-1} X'e\right] \tag{508}$$

Since $Rb - r = 0$

$$E(b) - \beta = M(X'X)^{-1} X'e \tag{509}$$

Where M is the idempotent matrix:

$$M = I - (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1} R \tag{510}$$

The variance covariance matrix of

$$cov(b) = [E(b) - \beta][E(b) - \beta]' = E\left[M(X'X)^{-1} X'ee'X(X'X)^{-1} M'\right] \tag{511}$$

$$cov(b) = \sigma^2 M(X'X)^{-1} M \tag{512}$$

$$cov(b) = \sigma^2 M(X'X)^{-1} \tag{513}$$

$$M = \sigma^2 \left\{I - (X'X)^{-1} R' \left[R(X'X)^{-1} R'\right]^{-1}\right\} R \tag{514}$$

Thus the variance of the restricted least square estimator is smaller than the variance of the unrestricted least square estimator.

1.11.1 Instrumental variables

Normal equations with instrumental variables

$$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{t-1} + u_t \tag{515}$$

X_{t-1} as instrument for $\alpha_2 Y_{t-1}$

Normal equations for two explanatory variable case

$$\sum Y_t = \widehat{\alpha}_0 N + \widehat{\alpha}_1 \sum X_t + \widehat{\alpha}_2 \sum X_{t-1} \tag{516}$$

$$\sum X_t Y_t = \widehat{\alpha}_0 \sum X_t + \widehat{\alpha}_1 \sum X_t^2 + \widehat{\alpha}_2 \sum X_{t-1} Y_{t-1} \tag{517}$$

$$\sum X_{t-1} Y_t = \widehat{\alpha}_0 \sum X_{t-1} + \widehat{\alpha}_1 \sum X_t X_{t-1} + \widehat{\alpha}_2 \sum X_{t-1} Y_{t-1} \tag{518}$$

This is different than the normal equations when instruments were not used.

$$\sum Y_t = \widehat{\alpha}_0 N + \widehat{\alpha}_1 \sum X_t + \widehat{\alpha}_2 \sum Y_{t-1} \tag{519}$$

$$\sum X_t Y_t = \widehat{\alpha}_0 \sum X_t + \widehat{\alpha}_1 \sum X_t^2 + \widehat{\alpha}_2 \sum X_{t-1} Y_{t-1} \tag{520}$$

$$\sum Y_{t-1} Y_t = \widehat{\alpha}_0 \sum Y_{t-1} + \widehat{\alpha}_1 \sum X_t Y_{t-1} + \widehat{\alpha}_2 \sum Y_{t-1}^2 \tag{521}$$

1.11.2 Sargan test (SARG) is used for validity of instruments

- Divide variables which are uncorrelated and correlated with the error terms $X_1, X_2, ..., X_p$ and $Z_1, Z_2, ..., Z_s$. Use instruments.$W_1, W_2, ..., W_p$
- Obtain estimates of $\widehat{u}_t$ from the original regression.
- Replace $Z_1, Z_2, ..., Z_s$ by instruments, $W_1, W_2, ..., W_p$.
- Regress on all and but exclude . Obtain R^2 of the regression.
- Compute SARG statistics $SARG = (n-k)\,R^2$ where n is the number of observations and k is the number of coefficients; SARG follows χ^2 distribution with $df = s - p$.
- H_0: W instruments are valid if the computed SARG exceed the χ^2critical value ; if H_0: is rejected at least one instrument is not valid.

Chesher A. and A. Rosen, "An instrumental variable random coefficients model for binary outcomes," forthcoming, Econometrics Journal

Chesher A., A. Rosen and K. Smolinski (2013) "An instrumental variable model for multiple discrete choice," Quantitative Economics, 4, 157-196.

1.12 Distributed Lag Models : Koyck, Almon, ARDL

$$\begin{aligned} C_t &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-3} + ... + +\beta_k X_{t-k} + \varepsilon_t \\ t &= 1 \,...T \end{aligned} \tag{522}$$

Reasons for lags

- Psychological reasons: it takes time to believe something.
- Technological reasons: takes time to change new machines or to update.
- Institutional reasons: rules, regulations, notices, contracts.

Lagged marginal effect in consumption of an increase in income at period 0.

Koyck's Model

- short run multiplier : β_1
- intermediate run multiplier: $\beta_1 + \beta_2 + \beta_3$
- long run multiplier: $\sum \beta_1 + \beta_2 + \beta_3 + ... + \beta_k$
- proportion of the long run impact at a certain period: $\beta^* = \frac{\beta_1}{\beta}$

Koyck's procedure: $\beta_2 = \lambda\beta_1; \beta_3 = \lambda^2\beta_1;\ \beta_k = \lambda^k\beta_1$ and so on.

$$C_t = \beta_0 + \beta_1 X_t + \lambda\beta_1 X_{t-1} + \lambda^2\beta_1 X_{t-2} + \lambda^3\beta_1 X_{t-3} + ... + \lambda^k\beta_1 X_{t-k} + \varepsilon_t \tag{523}$$

Koyck's procedure

Koyck procedure converts distributed lag model into an autoregressive model. It involves (a) multiplying (2) by λ, which is between 0 and 1, $0 < \lambda < 1$; (b) takking one period lag of that and (c) subtracting from (2)

$$\begin{aligned} \lambda C_t &= \lambda\beta_0 + \lambda\beta_1 X_t + \lambda^2\beta_1 X_{t-1} + \lambda^3\beta_1 X_{t-2} \\ &\quad + \lambda^4\beta_1 X_{t-3} + ... + \lambda^{k+1}\beta_1 X_{t-k} + \varepsilon_t \end{aligned} \tag{524}$$

$$\begin{aligned} C_t &= \beta_0 + \beta_1 X_t + \lambda\beta_1 X_{t-1} + \lambda^2\beta_1 X_{t-2} + \lambda^3\beta_1 X_{t-3} \\ &+ ... + \lambda^k\beta_1 X_{t-k} + \varepsilon_t \end{aligned} \tag{525}$$

$$\begin{aligned} \lambda C_{t-1} &= \lambda\beta_0 + \lambda\beta_1 X_{t-1} + \lambda^2\beta_1 X_{t-2} + \lambda^3\beta_1 X_{t-3} + \lambda^4\beta_1 X_{t-4} \\ &+ ... + \lambda^{k+1}\beta_1 X_{t-k-1} + \varepsilon_{t-1} \end{aligned} \tag{526}$$

Take the difference between these two

$$C_t - \lambda C_{t-1} = (1-\lambda)\beta_0 + \beta_1 X_t + \lambda^k\beta_1 X_{t-k} + \varepsilon_t \ - \varepsilon_{t-1} \tag{527}$$

Term $\lambda^k\beta_1 X_{t-k} \longrightarrow 0$ as $0 < \lambda < 1$

$$C_t = (1-\lambda)\beta_0 + \beta_1 X_t + \lambda C_{t-1} + u_t \tag{528}$$

$u_t = \varepsilon_t \ - \varepsilon_{t-1}$
By cancelling terms it transforms to an autoregressive equation as following:

In steady state $C_t = C_{t-1} = C$;

$$C = \beta_0 + \frac{\beta_1}{(1-\lambda)} X_t + \frac{u_t}{(1-\lambda)} \tag{529}$$

term $\frac{\beta_1}{(1-\lambda)}$ gives the long run impact of the change in X_t on C_t
Choice of Length of Lag in Koyck Model
Median lag: -$\frac{\log 2}{\log(\lambda)}$: 50% of the long run impact is felt over this lag
Mean lag: $\frac{\sum_{k=0}^{\infty} k\beta_k}{\sum_{k=0}^{\infty} \beta_k}$: mean of the total impact
Koyck mean lag: $\frac{\lambda}{(1-\lambda)}$: average lag length
How to choose lag length
Minimise Akaiki information criteria

$$AIC = \ln \frac{SSE_N}{T-N} + \frac{2(n+2)}{T-N} \tag{530}$$

Minimise Swartz criteria (minimise these values)

$$SC(N) = \ln \frac{SSE_N}{T-N} + \frac{2(n+2)\ln(T-N)}{T-N} \tag{531}$$

Problems with Koyck Model
• It is very restrictive. The successive coefficient may not decline geometrically when $0 < \lambda < 1$.
• There is no a-priori guide to the maximum length of lag ; Tinbergen suggests to use trial and error, first regress C_t on X_t and X_{t-1}, if the coefficients are significant, keep introducing lagged terms of higher order.
• But more lags implies fewer degrees of freedom

• Multicollinearity may appear
• Data mining
• Autoregressive term is correlated with the error term, Durbin-Watson statistics cannot be applied in this case. Need to use Durbin-h statistics which is defined as

$$h = \left(1 - \frac{1}{d}\right)\sqrt{\frac{T-1}{1-(T-1)\,SE\left(\beta_2\right)^2}} \tag{532}$$

Almon's polynomial lag model

Koyck procedure is very restrictive where the values of coefficients decline in geometric proportions. However impact of economic variables may be better explained by a quadratic cubic or higher order polynomial of the form:

$$C_t = \beta_0 + \beta_1 X_t + \lambda\beta_1 X_{t-1} + \lambda^2\beta_1 X_{t-2} + \lambda^3\beta_1 X_{t-3} + ... + \lambda^k\beta_1 X_{t-k} + \varepsilon_t \tag{533}$$

quadratic impact structure: $\beta_i = \alpha_0 + \alpha_1 \cdot i + \alpha_2 \cdot i^2 + \alpha_3$
cubic impact structure: $\beta_i = \alpha_0 + \alpha_1 \cdot i + \alpha_2 \cdot i^2 + \alpha_3 \cdot i^3$
k-order polynomial lag structure: $\beta_i = \alpha_0 + \alpha_1 \cdot i + \alpha_2 \cdot i^2 + \alpha_3 \cdot i^3 + ... + \alpha_k \cdot i^k$

$$C_t = \beta_0 + \frac{\beta_1}{(1-\lambda)}X_t + \frac{u_t}{(1-\lambda)} \tag{534}$$

$$C_t = \beta_0 + \sum_{k=0}^{\infty}\left(\alpha_0 + \alpha_1 \cdot i + \alpha_2 \cdot i^2 + \alpha_3 \cdot i^3 + ... + \alpha_k \cdot i^k\right)X_{t-1} + u_t \tag{535}$$

Advantages of Almon model over Koyck
a. Flexible; can incorporate variety of lag
b. do not have to decline geometrically, Koyck had rigid lag structure
c. No lagged dependent variable in the regression
d. Number of coefficient estimated significantly smaller than in the Koyck model
e. is likely to be multicollinear.

Estimates of a polynomial distributed lag model

Autoreggressive Distributed Lag Model: ARDL (1,1)

$$Y_t = \mu + \beta_0 X_t + \beta_1 X_{t-1} + \gamma Y_{t-1} + \varepsilon_t \tag{536}$$

This can be represented by an infinitely distributed lag as following

$$Y_t = \mu + \beta_0 X_t + \beta_1 X_{t-1} + \sum_{i=0}^{l}\gamma^{i-l}\left(\beta_1 + \gamma\beta_0\right)X_{t-1} + \varepsilon_t \tag{537}$$

lag weights:
$\alpha_0 = \beta_0$; $\alpha_1 = \left(\beta_1 + \gamma\beta_0\right)$; $\alpha_2 = \gamma\left(\beta_1 + \gamma\beta_0\right) = \gamma^2\alpha_1$; ....,$\alpha_S = \gamma^S\alpha_1$
ARDL (2,2)

$$Y_t = \mu + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \varepsilon_t \tag{538}$$

1.13 Simultaneous equation system

- Klein L. R. (1946) Macroeconomics and the Theory of Rational Behavior, Econometrica, 14, 2, 93-108
- Klein L. R. (1947) The Use of Econometric Models as a Guide to Economic Policy,Econometrica, 15, 2 , 111-151
- Klein L. R. (1971) Whither Econometrics?,Journal of the American Statistical Association, 66, 334, 415-421

Main Features of Simultaneous Equation System

- Single equation models have Y dependent variables to be determined by a set of X variables and the error term.
- one way causation from independent variables to the dependent variables.
- However, many variables in economics are interdependent and there is two way causation.
- Consider a market model with demand and supply.
- Price determines quantity and quantity determines price.
- Same is true in national income determination model. Consumption and income.

Main Feature of Simultaneous Equation System

- Both quantities and prices and income and consumption are determined simultaneously.
- A system of equations, not a single equation, need to be estimated in order to be able to capture this interdependency among variables.
- The main features of a simultaneous equation model are:

  (i) two or more dependent (endogenous) variables; a number of exogenous variables

  (ii) a set of equations
- Computationally cumbersome, highly non-linearity in parameters and errors in one equation transmitted through the whole system

Indentification issue in a Market Model

- Consider a relation between quantity and price

$$Q_t = \alpha_0 + \alpha_1 P_t + u_t \tag{539}$$

- A priory it is impossible to say whether this is a demand or supply model, both of them have same variables.
- If we estimate a regression model like this how can we be sure whether the parameters belong to a demand or supply model?
- We need extra information. Economic theory suggests that demand is related with income of individual and supply may respond to cost or weather condition; e.g. lagged price level P_{t-1}.

1.13.1 Market Model

In equilibrium quantity demand equals quantity supplied

$Q_t^d = Q_t^s$

$$\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t} = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t} \tag{540}$$

Solve for P_t

$$\alpha_1 P_t - \beta_1 P_t = \beta_0 - \alpha_0 + \beta_2 P_{t-1} + \alpha_2 I_t + u_{2,t} - u_{1,t} \tag{541}$$

$$P_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} I_t + \frac{\beta_2}{\alpha_1 - \beta_1} P_{t-1} + \frac{u_{2,t} - u_{1,t}}{\alpha_1 - \beta_1} \tag{542}$$

Using this price to solve for quantity

$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t} = \alpha_0 + \alpha_1 \left[\frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} I_t + \frac{\beta_2}{\alpha_1 - \beta_1} P_{t-1} + \frac{u_{2,t} - u_{1,t}}{\alpha_1 - \beta_1} \right] + \alpha_2 I_t + u_{1,t}$

$$Q_t = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} - \frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} I_t + \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1} P_{t-1} + \frac{\alpha_1 u_{2,t} - \beta_1 u_{1,t}}{\alpha_1 - \beta_1} \tag{543}$$

$$P_t = \Pi_{1,0} + \Pi_{1,1} P_{t-1} + \Pi_{1,2} I_t + V_{1,t} \tag{544}$$

$$Q_t = \Pi_{1,0} + \Pi_{1,1} P_{t-1} + \Pi_{1,2} I_t + V_{1,t} \tag{545}$$

$$\begin{aligned} \Pi_{1,0} &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \quad \Pi_{1,1} = \frac{-\alpha_2}{\alpha_1 - \beta_1} \; \Pi_{1,2} = \frac{-\beta_2}{\alpha_1 - \beta_1} \; ; V_{1,t} = \frac{u_{2,t} - u_{1,t}}{\alpha_1 - \beta_1} \\ \Pi_{2,0} &= \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} \; \Pi_{2,1} = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1} \qquad \Pi_{2,2} = \frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}; \end{aligned} \tag{546}$$

$$V_{1,t} = \frac{u_{2,t} - u_{1,t}}{\alpha_1 - \beta_1}; V_{2,t} = \frac{\alpha_1 u_{2,t} - \beta_1 u_{1,t}}{\alpha_1 - \beta_1} \tag{547}$$

1.13.2 Keynesian Model

$$C_t = \beta_0 + \beta_1 Y_t + u_t \tag{548}$$

$$Y_t = C_t + I_t \tag{549}$$

β_0 and $\beta_1$are structural parameters ; Y_t and C_t are endogenous variables and I_t is exogenous variable.

In the income determination model (example 2) the reduced form is obtained by expressing C and Y endogenous variables in terms of I which is the only exogenous variable in the model.

$$C_t = \frac{\beta_0}{1 - \beta_1} + \frac{\beta_1}{1 - \beta_1} I_t + \frac{1}{1 - \beta_1} u_{1,t} \tag{550}$$

$$Y_t = \frac{\beta_0}{1 - \beta_1} + \frac{1}{1 - \beta_1} I_t + \frac{1}{1 - \beta_1} u_{1,t} \tag{551}$$

$$C_t = \Pi_{1,1} + \Pi_{1,2} I_t + V_{1,t} \tag{552}$$

$$Y_t = \Pi_{2,1} + \Pi_{2,2} I_t + V_{2,t} \tag{553}$$

Retrieving the structural parameters of the model:

$$\Pi_{1,1} = \frac{\beta_0}{\beta_1} \quad \Pi_{1,2} = \frac{\beta_1}{1-\beta_1} \quad \Pi_{2,1} = \frac{\beta_0}{1-\beta_1} \quad \Pi_{2,2} = \frac{1}{1-\beta_1} \tag{554}$$

Keynesian Model: Simultaneity Bias

$$\widehat{\beta}_1 = \frac{\sum c_t y_t}{\sum y_t^2} = \frac{\sum \left(C_t - \overline{C}\right) y_t}{\sum y_t^2} = \frac{\sum C_t y_t}{\sum y_t^2} \tag{555}$$

$$\widehat{\beta}_1 = \frac{\sum C_t y_t}{\sum y_t^2} = \frac{\sum \left(\beta_0 + \beta_1 Y_t + u_t\ \right) y_t}{\sum y_t^2} \tag{556}$$

$$cov(Y,e) = E\left(Y_t - E\left(Y_t\right)\right) E\left(u_t - E\left(u_t\right)\right) = E\left(\frac{u_t}{1-\beta_1}\right) u_t = \frac{\sigma_e^2}{1-\beta_1} \tag{557}$$

$$p\lim\left(\widehat{\beta}_1\right) = \beta_1 + \frac{\sum u_t\ y_t}{\sum y_t^2} = \beta_1 + \frac{\frac{\sum u_t\ y_t}{T}}{\frac{\sum y_t^2}{T}} = \beta_1 + \frac{\frac{\sigma_e^2}{1-\beta_1}}{\sigma_y^2} \tag{558}$$

Techniques of estimation of simultaneous equation models

- Single Equations Methods: Recursive OLS
- Ordinary Least Squares
- Indirect Least Squares
- Two Stage Least Squares Method
- System Method
- Generalised Least Square
- Seemingly Unrelated Regression Equations

1.14 Recursive and 2SLS estimation

Recursive estimation

$$Y_{1,t} = \beta_{10} + \gamma_{11} X_{1,t} + \gamma_{12} X_{2,t} + e_{1,t} \tag{559}$$

$$Y_{2,t} = \beta_{20} + \beta_{21} Y_{1,t} + \gamma_{21} X_{1,t} + \gamma_{22} X_{2,t} + e_{2,t} \tag{560}$$

$$Y_{3,t} = \beta_{30} + \beta_{31} Y_{1,t} + \beta_{33} Y_{2,t} + \gamma_{31} X_{1,t} + \gamma_{32} X_{2,t} + e_{3,t} \tag{561}$$

Apply OLS to (1) and get the predicted value of $\widehat{Y}_{1,t}$. Then use $\widehat{Y}_{1,t}$ into equation (2) and apply OLS to equation (2) to get predicted value of $\widehat{Y}_{2,t}$. And Finally use predicted values of $\widehat{Y}_{1,t}$ and $\widehat{Y}_{2,t}$ to estimate in equation (3).

Two Stage Least Square Estimation (2SLS)

Consider a hybrid of Keynesian and classical model in which income $Y_{1,t}$ is function of money $Y_{2,t}$ investment $X_{1,t}$ and government spending $X_{2,t}$.

$$Y_{1,t} = \beta_{1,0} + \beta_{11} Y_{2,t} + \gamma_{11} X_{1,t} + \gamma_{12} X_{2,t} + e_{1,t} \tag{562}$$

$$Y_{2,t} = \beta_{2,0} + \beta_{21} Y_{1,t} + e_{2,t} \tag{563}$$

First estimate $Y_{1,t}$ is all exogenous variables.

$$Y_{1,t} = \widehat{\Pi}_{1,0} + \widehat{\Pi}_{1,1} X_{1,t} + \widehat{\Pi}_{1,2} X_{2,t} + \widehat{e}_{1,t} \tag{564}$$

Obtain predicted $\widehat{Y}_{1,t}$

$$\widehat{Y}_{1,t} = \widehat{\Pi}_{1,0} + \widehat{\Pi}_{1,1} X_{1,t} + \widehat{\Pi}_{1,2} X_{2,t} \tag{565}$$

Then put in

$$Y_{1,t} = \widehat{Y}_{1,t} + \widehat{e}_{1,t} \tag{566}$$

In the second stage put this into the money supply equation $Y_{2,t} = \beta_{2,0} + \beta_{21} Y_{1,t} + e_{2,t}$

$$Y_{2,t} = \beta_{2,0} + \beta_{21} \left(\widehat{Y}_{1,t} + \widehat{e}_{1,t} \right) + e_{2,t} \tag{567}$$

$$Y_{2,t} = \beta_{2,0} + \beta_{21} \widehat{Y}_{1,t} + \beta_{21} \widehat{e}_{1,t} + e_{2,t} \tag{568}$$

$$Y_{2,t} = \beta_{2,0} + \beta_{21} \widehat{Y}_{1,t} + e^*_{2,t} \tag{569}$$

$$e^*_{2,t} = \beta_{21} \widehat{e}_{1,t} + e_{2,t} \tag{570}$$

Application of the OLS in this equation gives consistent estimators.

Instrumental variable Method

Rank and Order Conditions for Identification

Order condition:

$$K - k \geqslant m - 1 \tag{571}$$

Rank condition: =>

$$\rho(A) \geqslant (M-1)(M-1) \tag{572}$$

order of the matrix.

M = number of endogenous variables in the model

K = number of exogenous variables in the model including the intercept

m = number of endogenous variable in an equation

k = number of exogenous variables in a given equation

Rank condition is defined by the rank of the matrix, which should have a dimension (M-1), where M is the number of endogenous variables in the model.

Determining the Rank of the Matrix

- 1. Rank of matrix is the order of non-singular matrix
  2. Rank matrix is formed from the coefficients of the variables (both endogenous and exogenous) excluded from that particular equation but included in the other equations in the model.
  3. The rank condition tells us whether the equation under consideration is identified or not.
  4. The order condition tells us if it is exactly identified or overidentified.

Steps for Rank Condition

- 1. Write down the system in the tabular form
  2. Strike out the coefficients corresponding to the equation to be identified
  3. Strike out the columns corresponding to those coefficients in 2 which are nonzero.
  4. The entries left in the table will give only the coefficients of the variables included in the system but not in the equation under consideration. From these coefficients form all possible A matrices of order M-1 and obtain a corresponding determinant. If at least one of these determinants is non-zero then that equation is identified.

Summary of Order and Rank Conditions of Identification

- 1. If $(K-k) > (m-1)$ and the order of rank $\rho(A)$ is M-1 then the concerned equation is overidentified.
  2. If $(K-k) = (m-1)$ and the order of rank $\rho(A)$ is M-1 then the equation is exactly identified.
  3. If $(K-k) \geqslant (m-1)$ and he order of rank $\rho(A)$ is less than M-1 then the equation is underidentified.
  4. If $(K-k) < (m-1)$ the structural equation is unidentified. The the order of rank $\rho(A)$ is less M-1 in this case.

1.14.1 Empirical part: procedure in PcGive

- 1. construct data set in macroeocnomic variables (Y, C, I , G, T , X, M, MS, i, inflation, wage rate, exchange rate etc)
  2. save data in *.csv format; e.g. macro.csv
  3. Start GiveWin and PcGive and open data file
  4. choose multiple equation dynamic modelling
  5. determine endogenous and exogenous variables and run simultaneous equation using 3SLS or FIML
  6. Study coefficients
  7. Change policy variables and construct few scenarios

Homeowrk: construct reasonable small scale macro model from the data in macro.csv. Project values of exogenous variables; do forecasts.

1.14.2 Excercise 9

1. Suppose that you have a simple model of consumption and income as following
   Consumption function:

$$C_t = \beta_0 + \beta_1 Y_t + u_t \tag{573}$$

National income identity:

$$Y_t = C_t + I_t \tag{574}$$

1. (a) Use rank and order conditions to find whether the consumption function is identified in this model.
   (b) Write a reduced form for this system. Show how you could retrieve the structural coefficientsβ_0 and β_1 if you applied OLS to this reduced form.
   (c) Show that application of OLS to (1) generates a biased estimate of β_1.
   (d) What other method would you recommend to get an unbiased and best estimator for this model? Write steps to be followed until you get the structural coefficients β_0 and β_1.
   (e) Write a short note on how this model could be used to make a historical simulation of consumption and income series.

2. Consider a market model for a particular product.
Demand: $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t}$ (1)
Supply: $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t}$ (2)
Here Q_t^d is quantity demanded and Q_t^s is quantity supplied, P_t is the price of commodity,P_{t-1} is price lagged by one period, I_t is income of an individual, $u_{1,t}$ and $u_{2,t}$ are independently and identically distributed (iid) error terms with a zero mean and a constant variance.Q_t and P_t are endogenous variables and $P_{t-1}$and I_t are exogenous variables α_0,α_1,α_2,and β_0,β_1,$\beta_2$are six parameters defining the system.

1. (a) How can simultaneity bias occur if one tries to apply OLS to each of the above equations.
   (b) Use rank and order conditions to judge whether each of these two equations are over-, under- or exactly identified.
   (c) Write down the reduced form for this system.
   (d) How would you estimate the coefficients of the reduced form equations? Write down the estimator.
   (e) If equations are identified explain how you may retrieve the structural parameters α_0,α_1,α_2,and β_0,β_1,β_2, and from the coefficients of the reduced form equations.

1.14.3 Seemming Unrelated Regression (SUR) MODEL

Pooling Cross Section and Time Series: Seemming Unrelated Regression (SUR) MODEL

- SUR if formed by stacking models

$$Y_1 = X_1 \beta + e_1 \tag{575}$$

$$Y_2 = X_2 \beta + e_2 \tag{576}$$

$$... \tag{577}$$

$$Y_m = X_m \beta + e_m \tag{578}$$

There are m equations and T observations in the SURE system (in growth rate example we have 151 countries and 31 observations). They can be stacked into one large equation system as following.

There are m equations and T observations in the SURE system (in growth rate example we have 151 countries and 31 observations). They can be stacked into one large equation system as following.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & . & . & 0 \\ 0 & X_2 & . & . & 0 \\ . & . & X_3 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & X_m \end{bmatrix} \begin{matrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_m \end{matrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_m \end{bmatrix} \tag{579}$$

- Each Y_m and $e_m$has a dimension of T by 1 and X_m has T by K dimension and each β_m has K by 1 dimension. The covariance matrix of errors has TM by TM dimension.

Seemming Unrelated Regression (SUR) MODEL: Assumptions

- Mean of $e_{i,t}$ is zero for every value of , $E(e_{i,t}) = 0$
- variance of $e_{i,t}$ is constant for every ith observation, $var(e_{1t}) = \sigma_i^2$
- $cov(e_{i,t}, e_{i,s}) = 0$ for al t=s; this also means there is no autocorrelation
- All of the above assumptions are standard to the OLS assumptions.
- The major difference lies on assumption of contemporaneous correlation across the disturbance terms in above two models.
- $cov(e_{i,t}, e_{j,s}) = \sigma_{i,j}^2$ The systems are related due to errors.

Variance Covariance Structure in SUR MODEL

- Dimension of each of the $\sigma_{i,j}$, like that of the identity matrix I, is T by T, and reflects the variance covariance matrix of the stacked regression.

- The Kronnecker product $\Sigma \otimes I$is a short way of writing this covariance matrix.
- Σ is the variance covariance matrix
- $\otimes$ is the symbol for the Kronnecker product
- I is Identity Matrix with T×M by T×M dimension.

$$ee\prime = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_m \end{bmatrix} \begin{bmatrix} e_1 & e_2 & . & . & e_m \end{bmatrix} = \begin{bmatrix} e_1^2 & e_1e_2 & e_1e_3 & e_1e_4 & e_1e_5 \\ e_1e_2 & e_2^2 & e_2e_3 & e_2e_4 & e_2e_5 \\ e_1e_3 & e_2e_3 & e_3^2 & e_3e_4 & e_4e_5 \\ e_1e_4 & e_4e_2 & e_4e_3 & e_4^2 & e_4e_5 \\ e_1e_5 & e_5e_2 & e_5e_3 & e_5e_4 & e_5^2 \end{bmatrix} \tag{580}$$

$$E\left(ee\prime\right) = \begin{bmatrix} var\left(e_1\right) & cov\left(e_1e_2\right) & cov\left(e_1e_3\right) & cov\left(e_1e_4\right) & cov\left(e_1e_5\right) \\ cov\left(e_1e_2\right) & var\left(e_2\right) & cov\left(e_2e_3\right) & cov\left(e_2e_4\right) & cov\left(e_2e_5\right) \\ cov\left(e_1e_3\right) & cov\left(e_2e_3\right) & var\left(e_3\right) & cov\left(e_3e_4\right) & cov\left(e_4e_5\right) \\ cov\left(e_1e_4\right) & cov\left(e_4e_2\right) & cov\left(e_4e_3\right) & var\left(e4\right) & cov\left(e_4e_5\right) \\ cov\left(e_1e_5\right) & cov\left(e_5e_2\right) & cov\left(e_5e_3\right) & cov\left(e_5e_4\right) & var\left(e_5\right) \end{bmatrix} \tag{581}$$

$$E\left(ee\prime\right) = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} & \sigma_{3,5} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_4^2 & \sigma_{4,5} \\ \sigma_{5,1} & \sigma_{5,2} & \sigma_{5,3} & \sigma_{5,4} & \sigma_5^2 \end{bmatrix} = V = \Sigma \otimes I \tag{582}$$

Application of the OLS technique individual equations generates inconsistent results. Sure method aims to correct this problem by estimating all equations simultaneously.

The SURE method is essentially a generalised least square estimator. Note

$$V^{-1} = \Sigma^{-1} \otimes I \tag{583}$$

Aitken generalised least square

$$\widehat{\beta} = \left[X'V^{-1}X\right]^{-1} X'V^{-1}Y = \left[X'\left(\Sigma^{-1} \otimes I\right)X\right]^{-1} X'\left(\Sigma^{-1} \otimes I\right)Y \tag{584}$$

$$\widehat{\beta} = \begin{bmatrix} \sigma_{1,1}X_1^{'}X_1 & \sigma_{1,1}X_1^{'}X_2 & \sigma_{1,1}X_1^{'}X_3 & \sigma_{1,m}X_1^{'}X_m \\ \sigma_{2,1}X_2^{'}X_1 & \sigma_{2,2}X_2^{'}X_2 & \sigma_{2,3}X_2^{'}X_3 & \sigma_{2,m}X_2^{'}X_m \\ & & & \\ \sigma_{m,1}X_m^{'}X_1 & \sigma_{m,2}X_m^{'}X_2 & \sigma_{m,3}X_m^{'}X_3 & \sigma_{m,4}X_m^{'}X_m \end{bmatrix} \begin{bmatrix} \sum\sigma_{1,j}X_1^{'}Y_j \\ \\ \\ \sum\sigma_{m,j}X_m^{'}Y_j \end{bmatrix} \tag{585}$$

Steps for SUR Estimation

- Estimate each equation separately using the least square technique.
- Use the least square residuals from step 1 to estimate the error term.
- Use the estimates from the second step to estimate two equations jointly within a generalised least square framework. If m=2 the variance covariance matrix will be as given below.

Estimation of Seemming Unrelated Regression (SUR) by GLS

- $$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{pmatrix} \tag{586}$$

  Using a theorem in matrix algebra W can be decomposed into two parts as

$$P'P = \Omega^{-1} \tag{587}$$

Use this partition of Ω to transform the original model as

$$Y = X\beta + \varepsilon \tag{588}$$

$$\beta_{OLS} = (X'X)^{-1} (X'Y) \tag{589}$$

Estimation of Seemming Unrelated Regression (SUR) by GLS
Transform it to

$$P'Y = P'X\beta + P'\varepsilon \tag{590}$$

$$Y^* = X^*\beta + \varepsilon^* \tag{591}$$

$$\beta_{GLS} = (X'P'PX)^{-1} (X'P'PY) \tag{592}$$

In matrix notation

$$\beta_{GLS} = \left(X^{*\prime}\Omega^{-1}X^*\right)^{-1} \left(X^{*\prime}\Omega^{-1}Y^*\right) \tag{593}$$

Ω^{-1} is inverse of variance covariance matrix.

The GLS estimates are best, linear and unbiased estimators of the coefficients in the SURE system.

1.15 Panel Data Model

Panel Data

for $i = 1, \ldots . N$ countries and $t = 1, \ldots ., T$ years

Table 14: Structure of Panel Data

Dependent Variable	Explanatory Variable	Random Error
$y_{1,1}$	$x_{1,1}$	$e_{1,1}$
.	.	.
$y_{1,T}$	$x_{1,T}$	$e_{1,T}$
$y_{2,1}$	$x_{2,1}$	$e_{2,1}$
.	.	.
$y_{2,T}$	$x_{2,T}$	$e_{2,T}$
.	.	.
$y_{N,1}$	$x_{N,1}$	$e_{,1}$
.	.	.
$y_{2,T}$	$x_{2,T}$	$e_{2,T}$

Panel Data Model: Fixed Effects

$$y_{i,t} = \alpha_i + x_{i,t}\beta + e_{i,t} \qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right) \tag{594}$$

where parameter α_i picks up the fixed effects that differ among individuals,β is the vector of coefficients on explanatory variables. These parameters can be estimated by OLS when N is small but not when that is large but the model need to be transformed to the least square dummy variable method when N is too large.

$$\overline{y}_i = \alpha_i + \overline{x}_i\beta + e_i \qquad \overline{y}_i = T^{-1}\sum_i y_{i,t} \tag{595}$$

$$y_{i,t} - \overline{y}_i = \left(x_{i,t} - \overline{x}_i\right)\beta + \left(e_{i,t} - e_i\right) \tag{596}$$

fixed effect least square dummy variable estimator of β is

$$\beta_{FE} = \left(\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)'\right)^{-1}\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(y_{i,t} - \overline{y}_i\right)' \tag{597}$$

$$\alpha_i = \overline{y}_i - \overline{x}_i\beta_{FE} \tag{598}$$

Panel Data Model: Fixed Effect
fixed effect least square dummy variable estimator of β is

$$\beta_{FE} = \left(\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)'\right)^{-1}\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(y_{i,t} - \overline{y}_i\right)' \tag{599}$$

$$\alpha_i = \overline{y}_i - \overline{x}_i\beta_{FE} \tag{600}$$

These estimators are unbiased, consistent and efficient with corresponding covariance matrix given by:

$$cov\left(\beta_{FE}\right) = \sigma_e^2\left(\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)'\right)^{-1} \tag{601}$$

$$\sigma_e^2 = \frac{1}{N\left(T-1\right)}\sum_t^T\sum_i^N \left(y_{i,t} - \alpha_i - x_{i,t}\beta_{FE}\right) \tag{602}$$

Panel Data Model: Random Effect
Random effect models are more appropriate for analysing determinants of growth as

$$y_{i,t} = \mu + x_{i,t}\beta + \alpha_i + e_{i,t} \tag{603}$$

where $\alpha_i \sim IID\left(0, \sigma_\alpha^2\right)$ are individual specific random errors and $e_{i,t} \sim IID\left(0, \sigma_e^2\right)$ are remaining random errors.

$$\alpha_i\iota_T + e_i \qquad where\ \iota_T = (1, 1, ...1) \tag{604}$$

$$var\left(\alpha_i \iota_{T} + e_i\right) = \Omega = \sigma_\alpha^2 \iota_{T} \iota_{T}' + \sigma_e^2 I_T \tag{605}$$

Errors are correlated therefore this requires estimation by the Generalised Least Square estimator. Transform the model by pre-multiplying by Ω^{-1} where

$$\Omega^{-1} = \sigma_e^2 \left[I_T - \frac{\sigma_\alpha^2}{\sigma_e^2 + T\sigma_\alpha^2} \iota_{T} \iota_{T}' \right] \tag{606}$$

Panel Data Model: Random Effect

$$\begin{aligned} \beta_{GLS} \quad = \quad & \left(\sum_t^T \sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)' + \psi T \sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)' \right)^{-1} \\ & \left(\sum_t^T \sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(y_{i,t} - \overline{y}_i\right)' + \psi T \sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(y_{i,t} - \overline{y}_i\right)' \right) \end{aligned} \tag{607}$$

$$\Omega = \begin{bmatrix} \sigma_\alpha^2 + \sigma_e^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & . & . & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_e^2 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & . & . & \sigma_\alpha^2 + \sigma_e^2 \end{bmatrix} \tag{608}$$

$$\Omega^{-\frac{1}{2}} = \frac{1}{\sigma_e} \left[I_T - 1 - \frac{\sigma_e}{\sqrt{\sigma_e^2 + T\sigma_\alpha^2}} \right] \tag{609}$$

$$\beta_{GLS} = \sum_i \left(X'\Omega^{-1}X\right)^{-1} \sum_i \left(X'\Omega^{-1}Y\right) \tag{610}$$

Panel Data Model: GMM Estimator

generalised method of moments (GMM) as proposed by Hansen (1982).

$$y_{i,t} = \gamma y_{i,t-1}\beta + \alpha_i + e_{i,t} \qquad \gamma < 1 \tag{611}$$

which generates the following estimator

$$\gamma_{FE} = \frac{\sum_t^T \sum_i^N \left(y_{i,t} - \overline{y}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right)}{\sum_t^T \sum_i^N \left(y_{i,t} - \overline{y}_{i,t-1}\right)^2}; \qquad \overline{y}_i = T^{-1} \sum_i y_{i,t}; \, and \ \overline{y}_{i,-1} = T^{-1} \sum_i y_{i,t-1} \tag{612}$$

This is not asymptotically unbiased estimator:

$$\gamma_{FE} = \gamma + \frac{\left(\frac{1}{NT}\right)\sum_t^T\sum_i^N \left(e_{i,t} - \overline{e}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right)}{\left(\frac{1}{NT}\right)\sum_t^T\sum_i^N \left(y_{i,t} - \overline{y}_{i,-1}\right)^2} \tag{613}$$

$$p\lim_{N\to\infty}\left(\frac{1}{NT}\right)\sum_t^T\sum_i^N \left(e_{i,t} - \overline{e}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right) = -\frac{\sigma_e^2}{T^2}\frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2} \neq 0 \tag{614}$$

Panel Data Model: Instrumental Variables for GMM
Instrumental variable methods have been suggested to solve this inconsistency

$$\widehat{\gamma}_{IV} = \frac{\sum_t^T\sum_i^N y_{i,t-2}\left(y_{i,t-1} - \overline{y}_{i,t-2}\right)}{\sum_t^T\sum_i^N y_{i,t-2}\left(y_{i,t-1} - y_{i,t-2}\right)} \tag{615}$$

where $y_{i,t-2}$ is used as instrument of $(y_{i,t-1} - y_{i,t-2})$
It is asymptotically

$$p\lim_{N\to\infty}\left(\frac{1}{NT}\right)\sum_t^T\sum_i^N \left(e_{i,t} - \overline{e}_i\right) y_{i,t-2} \tag{616}$$

Moment conditions with vector of transformed error terms

$$\Delta e_i = \begin{pmatrix} e_{i,2} - e_{i,1} \\ e_{i,3} - e_{i,2} \\ . \\ . \\ e_{i,T} - e_{i,T-1} \end{pmatrix} \tag{617}$$

Panel Data Model: Instrumental Variables for GMM 2

$$Z_i = \begin{bmatrix} [y_{i,0}] & 0 & . & . & 0 \\ 0 & [y_{i,0}, y_{i,1,}] & 0 & . & . \\ 0 & 0 & . & . & 0 \\ . & . & . & . & 0 \\ 0 & 0 & . & . & [y_{i,0}, y_{i,T-2}] \end{bmatrix} \tag{618}$$

$$E\left\{Z_i^{'}\Delta e_i\right\} = 0 \tag{619}$$

Or for moment estimator write the transformed errors as

$$E\left\{Z_i^{'}\left(\Delta y_{i,t} - \gamma\Delta y_{i,t}\right)\right\} = 0 \tag{620}$$

$$\min_{\gamma}\left(\left(\frac{1}{N}\right)\sum_{i=1}^N Z_i^{'}\left(\Delta y_{i,t} - \gamma\Delta y_{i,t}\right)\right)^{'} W_N\sum_{i=1}^N Z_i^{'}\left(\Delta y_{i,t} - \gamma\Delta y_{i,t}\right)^{'} \tag{621}$$

Panel Data Model: Instrumental Variables for GMM 2
GMM method includes the most efficient instrument

$$\gamma_{_{GMM}} = \left(\left(\sum_{i=1}^{N}\Delta y_{i,t}Z_i\right)W_N\left(\sum_{i=1}^{N}Z_i^{'}\Delta y_{i,t}\right)\right)^{-1}\times\left(\left(\sum_{i=1}^{N}\Delta y_{i,t}Z_i\right)W_N\left(\sum_{i=1}^{N}Z_i^{'}\Delta y_{i,t}\right)\right) \quad (622)$$

Blundell and Smith (1989) and Verbeek (2004), Wooldridge (2002) among others have more extensive exposure in GMM estimation. The essence of the GMM estimation remains in finding a weighting matrix that can guarantee the most efficient estimator. This should be inversely proportional to transformed covariance matrix.

$$W_N^{opt} = \left(\left(\frac{1}{N}\right)\sum_{i=1}^{N}Z_i^{'}\Delta e_{i,t}\Delta e_{i,t}^{'}Z_i\right)^{-1} \quad (623)$$

Panel Data Model: Instrumental Variables for GMM 2
Doornik and Hendry (2001, chap. 7-10) provide a procedure on how to estimate coefficients using fixed effect, random effect and the GMM methods including a lagged terms of dependent variable among explanatory variables for a dynamic panel data model:

$$y_{i,t} = \sum_{i=1}^{p}a_k y_{i,t-s} + \beta^t(L)\,x_{i,t} + \lambda_t + \alpha_i + e_{i,t} \quad or\ inshorty_{i,t} = W_i\delta + \iota_i a_i + e_i \quad (624)$$

The GMM estimator with instrument (levels, first differences, orthogonal deviations, deviations from individual means, combination of first differences and levels) used in PcGive is :

$$\widehat{\delta} = \left(\left(\sum_{i=1}^{N}W_i^{*}Z_i\right)A_N\left(\sum_{i=1}^{N}Z_i^{'}W_i\right)\right)^{-1}\left(\left(\sum_{i=1}^{N}W_i^{*}Z_i\right)A_N\left(\sum_{i=1}^{N}Z_i^{'}y_i^{*}\right)\right) \quad (625)$$

where $A_N = \left(\sum_{i=1}^{N}Z_i^{'}H_iZ_i\right)^{-1}$ is the individual specific weighting matrix.
Panel Estimation

Panel Cointegration

1.15.1 Panel Cointegration

Long run relationship obtained in the dynamic general equilibrium are tested by the GMM estimation of dynamic panel model. The determinants of growth of per capita output and the exchange rates across eleven countries representing the global economy in fact validate the conclusion of general equilibrium results. Estimates support the standard neoclassical theory of economic growth and uncovered interest parity theory of exchange rate though country specific factors, including preferences and technology, can also have significant influence in estimation of each of these models.

Table 15: Determinants of growth rate of per capita income and Exchange Rate

	Growth Model		Exchange Rate Model	
Determinants	Coefficient	t-prob	Coefficient	t-value
Investment ratio	0.1820	.00060	-	-
Export Ratio	0.0257	.3830	-	-
Exchange rate -1	-	-	0.9710	0.00
Real Interest rate	-	-	-0.0290	0.00
Population growth rate	-0.8849	0.1540	0.7917	0.00
Constant	3.0116	0.1780	0.3400	0.00
Nepal	-3.0341	0.0000	0.0662	0.00
India	-2.0244	0.0000	0.0496	0.00
South Africa	-5.1070	0.0000	0.0709	0.00
Brazil	-4.5529	0.0000	-0.0324	0.00
UK	-4.5630	0.0020	0.0031	0.00
Japan	-5.9846	0.0000	-0.0422	0.00
USA	-3.7902	0.0000	0.0295	0.00
Germany	-5.6408	0.0000	-0.0074	0.00
	N =324	R^2= 0.46	N =312	R^2= 0.9857

1.15.2 Exercise 10

Consider the cross-regional variation of expenditure on food in the UK. For simplicity, it is assumed that food expenditure depends only on wage and salary income in each region.

1. (a) Formulate a model relating expenditure on food (F) and income (Y) that takes account of region specific effects. Note that the equations for each region are independent but that there is contemporaneous correlation among the error terms across the regions. State the major assumptions of the model.

   (b) Represent the model in terms of a system of stacked regressions that takes account of both individual and system specific effects. What is the structure of the covariance matrix of the error terms in this system?

   (c) Show how the SURE or GLS estimator system can be applied to estimate the structural parameters of this model. Write out their covariance structure in the matrix form.

   (d) This model has been estimated using a pooled time series and cross section data set (with the sample size of T=14 and N=13) available from the web site of the Office of the National Statistics (hhttp://www.statistics.gov.uk). The estimated coefficients, by region, are given in the following table. Analyse and interpret these results.

Similar models could be counstructed to study demand for utilities -electricity, water, telephone,

transport; or regional variation in growth, investment, research and development, employment, wage, credit flows, production and consumption of several agricultural products.

2. Consider a panel data regression model aimed to measure the impacts of FDI on economic growth as following:

Table 16: Cointegration Test of Growth and Exchange Rate Equations

	Cointegration in Growth Model		Cointegration in Exchange rate Model	
ADF test (T=321; Constant; 5%=-2.87; 1% = -3.45)				
Determinants	ADF-Statistics	Decision	ADF-Statistics	Decision
Investment ratio	-4.449**	Stationary	-	-
Export Ratio	-1.9000	Non-Stationary	-	-
Exchange rate -1	-	-	-1.510	Non-Stationary
Real Interest rate	-	-	-2.59	Non-Stationary
Population growth rate	-6.171**	Stationary	-6.171**	Stationary
Redidual	-10.62**	Stationary	-4.96**	Stationary
Conclusion: Variables in both growth and exchange rates equations are cointegrated.				

$$y_{i,t} = \alpha_i + \beta_1 y_{i,t-1} + \beta_2 F_{i,t} + \beta_3 T_{i,t} + \beta_4 I_{i,t-1} + e_{i,t} \qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right) \tag{626}$$

where $y_{i,t}$ is the growth rate $F_{i,t}$ FDI ratio to GDP, $T_{i,t}$ is the ratio of tax revenue, $I_{i,t-1}$ is the ratio of investment. Use data in panel_fdi.csv to estimate this model using PcGive. Interprete your results.

3. Construct a panel data on growth rate of per capita income, investment ratio, population growth, export, imports, exchange rate, inflation rate for any five country of your choice. Suggest a panel growth model to be estimated.

Action:

Construct data on growth rate of per capita GDP, exchange rate and inflation rates from the www.imf.org for year 1980 to 2009 for China, India, South Africa, UK, USA and Brazil from the World Economic Outlook Database. Test whether inflation and the exchange rate are the significant variables in explaining the growth rate of per capita output (in PPP) in these economies using random or fixed effect models.

1.16 Linear probability, probit and logit models

- Alternative names: dichotomous dependent variables, discrete dependent random variable, binary variable, either or choice variables

$$Y_i = \int \begin{array}{ll} Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i & \text{if the event occurs} \\ 0 = \text{otherwise} & \end{array} \tag{627}$$

  Examples

- the labour force participation (1 if a person participates in the labour force, 0 otherwise)
- yes or no vote in particular issue ; to marry or not to marry; to study further or to start a job
- to buy or not to buy a particular stock
- choice of transportation mode to work (1 if a person drives to work, 0 otherwise)

- Union membership (1 if one is a member of the union, 0 otherwise)
- Owning a house (1 if one owns 0 otherwise)
- Multinomial choices: work as a teacher, or as a clerk, or as a self employed or professional or as a factory worker
- Multinomial ordered choices: strongly agree, agree, neutral, disagree

Linear Probability Model

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{628}$$

where $Y_i = 1$ if person owns a house, 0 otherwise; X_i is family income.
$E\left[\left(Y_i = 1\right)/X_i\right]$ probability that the event y will occur given x

$$E\left[\left(Y_i = 1\right)/X_i\right] = 0 \times \left[1 - P_i\right] + 1 \times P_i = P_i \tag{629}$$

$$0 \leqslant E\left[\left(Y_i = 1\right)/X_i\right] = P_i = \beta_1 + \beta_2 X_i \leqslant 1 \tag{630}$$

- Problem: Errors are heteroscedastic

$$\varepsilon_i = 1 - \beta_1 - \beta_2 X_i \quad with \ \ (1 - P_i) \tag{631}$$

$$\varepsilon_i = -\beta_1 - \beta_2 X_i \quad with \ \ P_i \tag{632}$$

Variance of error in a linear probability model

$$var\left(\varepsilon_i\right) = \left(1 - \beta_1 - \beta_2 X_i\right)^2 \ \left(1 - P_i\right) + \left(-\beta_1 - \beta_2 X_i\ \right)^2 P_i \tag{633}$$

$$\sigma^2 = \left(1 - \beta_1 - \beta_2 X_i\right)^2 \ \left(-\beta_1 - \beta_2 X_i\right) + \left(-\beta_1 - \beta_2 X_i\ \right)^2 \left(1 - \beta_1 - \beta_2 X_i\right) \tag{634}$$

$$\sigma^2 = \left(1 - \beta_1 - \beta_2 X_i\right)\left(\beta_1 + \beta_2 X_i\right) \ = \left(1 - P_i\right) \ P_i \tag{635}$$

Variance depends on X.
Limitations of a linear probability model
It is possible to transform this model to make it homescedastic by dividing the original variables by

$$\sqrt{\left(1 - \beta_1 - \beta_2 X_i\right)\left(\beta_1 + \beta_2 X_i\right)} \ = \sqrt{\left(1 - P_i\right) \ P_i} = \sqrt{W_i} \tag{636}$$

$$\frac{Y_i}{\sqrt{W_i}} = \frac{\beta_1}{\sqrt{W_i}} + \beta_2 \frac{X_i}{\sqrt{W_i}} + \frac{\varepsilon_i}{\sqrt{W_i}} \tag{637}$$

- It does not guarantee that the probability lies inside (0,1) bands

- Probability in non-linear phenomenon: at very low level of income a family does not own a house; at very high level of income every one owns a house ; marginal effect of income is very negligible. The linear probability model does not explain this fact well.

Probit Model

- $$\begin{aligned}\Pr(Y_i = 1) &= \Pr(Z_i^* \leq Z_i) = F(Z_i) = \frac{1}{\sqrt{2\pi}} \int \begin{array}{c} Z_i \\ -\infty \end{array} e^{-\frac{t^e}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int \begin{array}{c} \beta_1 + \beta_2 X_i + \varepsilon_i \\ -\infty \end{array} e^{-\frac{t^e}{2}} dt \end{aligned} \tag{638}$$

- Here t is standardised normal variable, $t \sim N(0,1)$

  probability depends upon unobserved utility index Z_i which depends upon observable variables such as income. There is a thresh-hold of this index when after which family starts owning a house, $Z_i \geqslant Z_i^*$

Logit Model

- variable Y_i which takes value 1 $(Y_i = 1)$ if a student gets a first class mark, value 0 $(Y_i = 0)$otherwise.
- Probability of getting a first class mark in an exam is a function of student effort index denoted by $.Z_i$; where $P_i = \frac{1}{1+e^{-Z_i}}$

  $Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ An example of a logit model: what determines that a student gets the first class degree?

$$Z_i = \beta_1 + \beta_2 H_i + \beta_3 E_i + \beta_4 A_i + \beta_2)_i + \varepsilon_i \tag{639}$$

  H = hours of study, E= exercises, A = attendance in lectures and classes; P = papers written for assignment.

- Ratio of odds: $\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i}$;taking log of the odds $ln\left(\frac{P_i}{1-P_i}\right) = Z_i$

Features of a logit Model

- probability goes from 0 to 1 as the index variable goes from -∞ to +∞. Probability lies between 0 and 1.
- Log of the odds is linear in x, characteristic variables but probabilities themselves are not linear but non linear function of the parameters. Probabilities are estimated using the maximum likelihood method.
- Any explanatory variable that determines the value of Z_i , measures how the log of odds of an event (i.e. owning a house) changes as a result of change in explanatory variable such as income.
- We can calculate P_i for given estimates of $\beta_1$and $\beta_2$or all other $.\beta_i$

- Limiting case when P_i =1; $ln\left(\frac{P_i}{1-1}\right)$ or when P_i =0 ; $ln\left(\frac{0}{1-0}\right)$OLS cannot be applied in such case but the maximum likelihood method may be used to estimate the parameters.

Logit model on probability of getting married from the dataset constructed from the BHPS (Hours.csv)

Table 17: Probability of Getting Married

	Coefficient	t-value	t-prob
Intercept	-2.99	-8.44	0.000
Log workhours	0.277	2.13	0.034
Gender	0.269	4.33	0.000
Labour	0.187	2.74	0.006
Liberal	0.330	3.28	0.001
Conservative	0.381	4.60	0.000
Health	0.189	6.56	0.000
Money	-0.036	-2.49	0.013
Children	0.253	23.5	0.000
Job	-0.124	-7.43	0.000
State =2 , AIC = 7244.8 N = 5790; LL -3612.4			

Tobit Model

It is an extension of the probit model, named after Tobin. We observe variables if the event occurs: ie if some one buys a house. We do not observe explanatory variables for people who have not bought a house. The observed sample is censored, contains observations for only those who buy the house.

$$Y_i = \int \begin{array}{ll} Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i & \text{if the event occurs} \\ 0 = \text{otherwise} & \end{array} \tag{640}$$

is equal to is the event is observed equal to zero if the event is not observed.

It is unscientific to estimate probability only with observed sample without worrying about the remaining observations in the truncated distribution. The Tobit model tries to correct this bias.

Inverse Mill's ratio: Example first estimate probability of work then estimate the hourly wage as a function of socio-economic background variables

Summary of Probability Models

The effect of observed variables on probability

$$\frac{\partial P_i}{\partial x_{i,j}} = \left\{ \begin{array}{l} \beta_j \\ \beta_j P_j \left(1 - P_j\right) \\ \beta_j \phi \left(Z_i\right) \end{array} \right. \tag{641}$$

where $Z_i = \beta_0 + \sum\limits_{i=1}^{k} \beta_i X_{i,j}$ and ϕ is the standard normal density function.

Estimate probability models using data in Hours.csv.

1.16.1 AR, MA, ARMA and ARIMA Forecasting

AR(1) forecast

$$y_t = \delta + \theta_1 y_{t-1} + e_t \tag{642}$$

h =1 ahead Forecast

$$y_{T+1} = \delta + \theta_1 y_T + e_{T+1} \qquad e_{T+1} \sim N(0,1) \tag{643}$$

Mean forecast:

$$\widehat{y}_{T+1} = E\left(y_{T+1}\right) = \delta + \theta_1 y_T \tag{644}$$

Estimate of Forecast error

$$\widehat{e}_{T+1} = y_{T+1} - \widehat{y}_{T+1} = \delta + \theta_1 y_T + e_{T+1} - \delta - \theta_1 \widehat{y}_T \tag{645}$$

Variance of h =1 Forecast error

$$var\left(\widehat{e}_{T+1}\right) = \sigma_e^2 \tag{646}$$

h =2 ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{T+1} + e_{T+1} \qquad e_{T+2} \sim N(0,1) \tag{647}$$

Mean forecast:

$$\widehat{y}_{T+2} = E\left(y_{T+2}\right) = \delta + \theta_1 y_{T+1} \tag{648}$$

Estimate of Forecast error

$$\begin{aligned} \widehat{e}_{T+2} &= y_{T+2} - \widehat{y}_{T+2} = \delta + \theta_1 y_{T+1} + e_{T+2} - \delta - \theta_1 \widehat{y}_{T+1} \\ &= e_{T+2} + \theta_1\left(y_{T+1} - \widehat{y}_{T+1}\right) = e_{T+2} + \theta_1 e_{T+1} \end{aligned} \tag{649}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+2}\right) = \sigma_e^2\left(1 + \theta_1^2\right) \tag{650}$$

h period ahead Forecast

$$y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h} \qquad e_{T+h} \sim N(0,1) \tag{651}$$

Mean forecast:

$$\widehat{y}_{T+h} = E\left(y_{T+h}\right) = \delta + \theta_1 \widehat{y}_{T+h-1} \tag{652}$$

Estimate of Forecast error

$$\begin{aligned} \widehat{e}_{T+h} &= y_{T+h} - \widehat{y}_{T+2} = \delta + \theta_1 y_{T+h-1} + e_{T+h} - \delta - \theta_1 \widehat{y}_{T+h-1} \\ &= e_{T+h} + \theta_1\left(y_{T+h-1} - \widehat{y}T+h-1\right) = e_{T+h} + \theta_1 e_{T+h-1} \end{aligned} \tag{653}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+h}\right) = \sigma_e^2\left(1 + \theta_1^2 + \theta_1^2 + ... + \theta_1^{2(h-1)}\right) \tag{654}$$

MA(1) forecast Forecast with MA(1)

$$y_t = \mu + e_t + \alpha_1 e_{t-1} \tag{655}$$

h=1 period ahead forecast

$$y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T \tag{656}$$

Mean forecast

$$E\left(y_{T+1}\right) = \widehat{y}_{T+1} = \mu + \alpha_1 e_T \tag{657}$$

Forecast error

$$y_{T+1} - \widehat{y}_{T+1} = \mu + e_{T+1} + \alpha_1 e_T - \mu - \alpha_1 e_{T+1} = e_{T+1} \tag{658}$$

Variance of forecast:

$$var\left(y_{T+1} - \widehat{y}_{T+1}\right) = var\left(e_{T+1}\right) = \sigma_e^2 \tag{659}$$

h=2 period ahead Forecast

$$y_{T+2} = \mu + e_{T+2} + \alpha_1 e_{T+1} \tag{660}$$

Mean forecast

$$E\left(y_{T+2}\right) = \widehat{y}_{T+2} = \mu \tag{661}$$

Forecast error

$$y_{T+2} - \widehat{y}_{T+2} = \mu + e_{T+2} + \alpha_1 e_{T+1} - \mu = e_{T+2} + \alpha_1 e_{T+1} \tag{662}$$

$$var\left(y_{T+2} - \widehat{y}_{T+2}\right) = var\left(e_{T+2}\right) = var\left(e_{T+2} + \alpha_1 e_{T+1}\right) = \sigma_e^2\left(1 + \alpha_1^2\right) \tag{663}$$

Similarly mean and variance of h period ahead forecast:

$$y_{T+h} = \mu + e_{T+h} + \alpha_1 e_{T+h-1} \tag{664}$$

$$E\left(y_{T+h}\right) = \widehat{y}_{T+h} = \mu \tag{665}$$

Forecast error

$$y_{T+h} - \widehat{y}_{T+h} = \mu + e_{T+h} + \alpha_1 e_{T+h-1} - \mu = e_{T+h} + \alpha_1 e_{T+h-1} \tag{666}$$

$$var\left(y_{T+2} - \widehat{y}_{T+2}\right) = var\left(e_{T+h}\right) = var\left(e_{T+h} + \alpha_1 e_{T+h-1}\right) = \sigma_e^2\left(1 + \alpha_1^2\right) \tag{667}$$

ARMA(1,1) forecast Forecasts using ARMA(1,1) process:

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1} \tag{668}$$

h=1 period ahead Forecast

$$y_{T+1} = \delta + \theta_1 y_{t-1} + e_{T+1} + \alpha_1 e_T \tag{669}$$

Mean forecast

$$E\left(y_{T+1}\right) = \widehat{y}_{T+1} = \delta + \theta_1 y_{t-1} + \alpha_1 e_T \tag{670}$$

Forecast error

$$\begin{aligned} \widehat{e}_{T+1} &= \left(y_{T+h} - \widehat{y}_{T+h}\right) = \\ \delta + \theta_1 y_{t-1} + e_{T+1} + e_{T+1} + \alpha_1 e_T - \delta - \theta_1 y_{t-1} - \alpha_1 e_T &= e_{T+1} \end{aligned} \tag{671}$$

Forecast error

$$\begin{aligned} \widehat{e}_{T+1} &= \left(y_{T+h} - \widehat{y}_{T+h}\right) = \delta + \theta_1 y_{t-1} + e_{T+1} + e_{T+1} \\ +\alpha_1 e_T - \delta - \theta_1 y_{t-1} - \alpha_1 e_T &= e_{T+1} \end{aligned} \tag{672}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+1}\right) = var\left(y_{T+h} - \widehat{y}_{T+h}\right) = var\left(e_{T+1}\right) = \sigma_e^2 \tag{673}$$

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1} \tag{674}$$

h=2 period ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1} \tag{675}$$

Mean forecast and Forecast error

$$E\left(y_{T+2}\right) = \widehat{y}_{T+2} = \delta + \theta_1 y_{t+1} \tag{676}$$

$$\begin{aligned} \widehat{e}_{T+2} &= \left(y_{T+2} - \widehat{y}_{T+2}\right) = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1} - \delta - \theta_1 \widehat{y}_{T+1} \\ &= \theta_1\left(y_{t+1} - \widehat{y}_{T+1}\right) + e_{T+2} + \alpha_1 e_{T+1} = \left(\theta_1 + \alpha_1\right) e_{T+1} + e_{T+2} \end{aligned} \tag{677}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+1}\right) = var\left[\left(\theta_1 + \alpha_1\right) e_{T+1} + e_{T+2}\right] = var\left(e_{T+1}\right) = \sigma_e^2\left[\left(\theta_1 + \alpha_1\right)^2 + 1\right] \tag{678}$$

h=3 period ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{t+2} + e_{T+3} + \alpha_1 e_{T+2} \tag{679}$$

Mean forecast

$$E\left(y_{T+3}\right)=\widehat{y}_{T+3}=\delta+\theta_1\widehat{y}_{t+2} \tag{680}$$

Forecast error and Variance of Forecast error

$$\begin{aligned}
\widehat{e}_{T+3} &= \left(y_{T+3}-\widehat{y}_{T+3}\right)=\delta+\theta_1 y_{t+2}+e_{T+3}+\alpha_1 e_{T+2}-\delta-\theta_1\widehat{y}_{T+2} \\
&= \theta_1\left(y_{t+2}-\widehat{y}_{T+2}\right)+e_{T+3}+\alpha_1 e_{T+2} \\
&= e_{T+3}+\alpha_1 e_{T+2}+\left(\theta_1+\alpha_1\right)e_{T+2}+e_{T+2}
\end{aligned} \tag{681}$$

$$\begin{aligned}
var\left(\widehat{e}_{T+3}\right) &= var\left[e_{T+3}+\alpha_1 e_{T+2}+\left(\theta_1+\alpha_1\right)e_{T+2}+e_{T+2}\right] \\
&= \sigma_e^2\left[1+\left(1+\alpha_1\right)^2+\left(\theta_1+\alpha_1\right)^2\right]
\end{aligned} \tag{682}$$

1.17 VAR Analysis

Consider a vector autoregressive model of order 2, VAR(2) given below.

$$y_t=a_{10}+a_{11}y_{t-1}+a_{12}y_{t-2}+b_{11}x_{t-1}+b_{12}x_{t-2}+e_{1,t} \tag{683}$$

$$x_t=a_{20}+a_{21}y_{t-1}+a_{22}y_{t-2}+b_{21}x_{t-1}+b_{22}x_{t-2}+e_{2,t} \tag{684}$$

where and are two variables for time t range from 1 ... T periods. Errors of each equation, $e_1$and e_2, are identically and independently distributed with zero mean and constant variance and covariance between and is assumed zero.

a. Evaluate the relationship between and in the long run.

Answer: Long run relationship is obtained by imposing the steady state relations:

$$\overline{y}=a_{10}+a_{11}\overline{y}+a_{12}\overline{y}+b_{11}\overline{x}+b_{12}\overline{x} \tag{685}$$

$$\overline{y}=\frac{a_{10}}{1-a_{11}-a_{12}}+\frac{\left(b_{11}+b_{12}\right)}{1-a_{11}-a_{12}}\overline{x} \tag{686}$$

$$\overline{x}=a_{20}+a_{21}\overline{y}+a_{22}\overline{y}+b_{21}\overline{x}+b_{22}\overline{x} \tag{687}$$

$$\overline{x}=\frac{a_{20}}{1-b_{21}-b_{22}}+\frac{\left(a_{21}+a_{22}\right)}{1-b_{21}-b_{22}}\overline{y} \tag{688}$$

b. Provide impulse response analysis for and of a unit shock in $e_{1,t}$ and $e_{2,t}$.

Use lag operator $y_{t-1}=Ly_t; y_{t-2}=Ly_{t-1}=L^2y_t$;Then the system changes to

$$y_t=a_{10}+a_{11}Ly_t+a_{12}L^2y_t+b_{11}Lx_t+b_{12}L^2x_t+e_{1,t} \tag{689}$$

$$x_t=a_{20}+a_{21}Ly_t+a_{22}L^2y_t+b_{21}Lx_t+b_{22}L^2x_t+e_{2,t} \tag{690}$$

$$y_t = \frac{a_{10}}{1 - a_{11}L - a_{12}L^2} + \frac{(b_{11} + b_{12})}{1 - a_{11}L - a_{12}L^2} x_t + \frac{1}{1 - a_{11}L - a_{12}L^2} e_{1,t} \tag{691}$$

$$x_t = \frac{a_{10}}{1 - b_{11}L - b_{12}L^2} + \frac{(a_{11} + a_{12})}{1 - b_{11}L - b_{12}L^2} y_t + \frac{1}{1 - b_{11}L - b_{12}L^2} e_{2,t} \tag{692}$$

Terms $\frac{1}{1-a_{11}L-a_{12}L^2}e_{1,t}$ and $\frac{1}{1-b_{11}L-b_{12}L^2}e_{2,t}$ give the impulse response of the first and second equations respectively.

c. Indicate and explain criteria to determine the order of a VAR model like this:

It is wise to use from general to specific approach of David Hendry to determine the order of VAR . First start the model with a large number of lags and then keep reducing the number of lags until the significant relation is found. Likelihood ratio tests are suggested for this.

d. What extra information is needed to make a h period ahead forecast using the above model? VAR is a time series model. Given the past values of time series, it requires distribution of the error terms for h period ahead forecasts.

e. A diagram can show how the variance of the forecast error and the confidence interval of a forecast are sensitive to the number of periods in the forecast horizon. The confidence level of forecast increases with the larger horizon of the forecasts.

1.17.1 ARCH/GARCH modelling of volatility

OLS estimates are based on the normality of errors, which are assumed to have constant mean and variance. Engle (1983) argued that many economic time series go through a series of ups and downs. Upward trend continues up to a significant length of time. and so does the downward trend.

As such the conditional mean and variance of these series are not constant. Modelling mean and vaiance of series simultaneously is the essence of the autoregressive conditional heteroskedasticity (ARCH) model.

The variance of error term is persistent and shown by autoregressive process of variances.

This technique has been widely used to measure the volatitiy of financial time siries such as the interest rate, inflation, stock prices, returns to assets, growth rates, trends in trades.

Bollersleve (1987) modified it to generalised autoregressive conditional heteroskedasticity (GARCH) models.

How ARCH and GARCH models are used to test the heteroskedasticity are discussed first followed by illustrations on variants of them used to study the clustering of heteroskedastic errors commonly used in the literature.

Engle (1983) autoregressive conditional heteroskedasticy (ARCH): more useful for time series data

Model $Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + ... + \beta_k X_{k,t} + e_t$

$\varepsilon_t \sim N\left(0, \left(\alpha_0 + \alpha_2 e_{t-1}^2\right)\right)$

$$\sigma_t^2 = \alpha_0 + \alpha_2 e_{t-1}^2 \tag{693}$$

Here σ_t^2 not observed. Simple way to estimate this is to run OLS of Y_t and get $\widehat{e}_t^2$. Then assume an ARCH (1) of errors as

$\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + v_t$ or ARCH (p) $\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{1-1}^2 + \alpha_4 \widehat{e}_{1-1}^2 + .. + \alpha_p \widehat{e}_{1-p}^2 + v_t$

Compute the test statistics

$n.R^2 \sim \chi_{df}^2$

Again if the calculated χ^2_{df} is greater than table value there is an evidence of ARCH effect and heteroskedasticity.

Economies are characterised by torbulent high volatility periods followed by quite and peaceful low volatity periods.

Decision makers require some estimates of expected values as well as volatility to reflect on the uncertainties causes by such phenomenon.

Recently stock prices rised contineously from 2002 to mid 2008 and then fell sharply in 2008 and 2009 and can be expected to rise in the next few years. Billions are lost and won because of volatilities in these series.

Engle (1987) proposes modelling expected value and volatility simultaneously by ARCH using iterative Maximum Likelihood procedure.as:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + e_t \tag{694}$$

where $e_t \sim N\left(0, \sigma_t^2\right) = N\left(0, h_t\right); \ h_t = \sigma_t^2.$

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 \tag{695}$$

Bollerslev (1987) generalised autoregressive conditional heteroskedasticy (GARCH) process is more general. For instance GARCH (1,1). Mean and variance equations take the following form:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + e_t \tag{696}$$

$$\sigma_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \beta \sigma_{t-1}^2 + v_t \tag{697}$$

GARCH (p,q)

$\sigma_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{t-2}^2 + \alpha_4 \widehat{e}_{t-3}^2 + .. + \alpha_p \widehat{e}_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + ..\beta_q \sigma_{t-q}^2 + .. + v_t$

Compute the test statistics $n.R^2 \sim \chi^2_{df}$

Sometimes written as

$h_t = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{t-2}^2 + \alpha_4 \widehat{e}_{t-3}^2 + .. + \alpha_p \widehat{e}_{t-p}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + ..\beta_q h_{t-q} + .. + v_t$

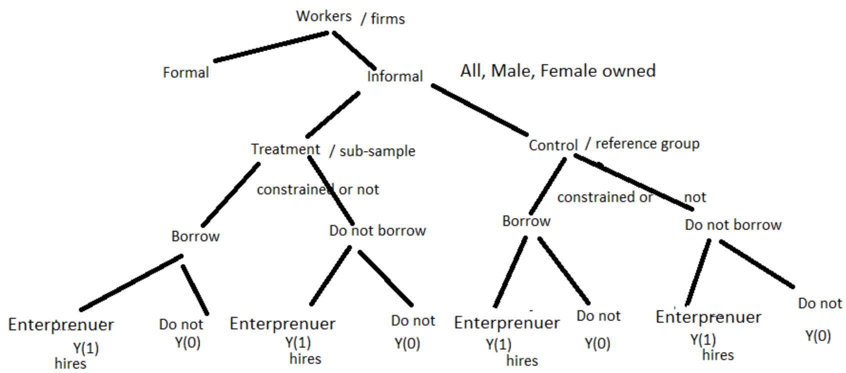
$where \quad h_t = \sigma_t^2$

1.17.2 Homework

1. Select one time series such as stock price or quarterly consumption. Estimate AR, MA, ARMA models to make ten period ahead forecasts. Use data in stocks.csv.

2. Estiamte a VAR(2) of growth rate of DPD and inflation rate. Do impulse response analysis using unit shocks to ten period horizon. Provide the forecast and the confidence interval of forecasst.

3. Take daily stock price of a certain company. Fit appropriate ARCH/GARCH models to explain volatility. Make a forecast.



1.18 Theory of causal inference: propensity score matching and difference in difference methods

Theory of causal inference: propensity score matching and difference in difference methods

Theory of causal inference: propensity score matching and difference in difference methods

- mathematics of treatment-effects estimators; Following Angrist, J. and Imbens, G., 1995. Identification and estimation of local average treatment effects:

- Let $Y_i(0)$ be the response without the treatment or program for individual i. $Y_i(1)$ is the response with treatment; We observe Di and $Yi = Y1(Di) = Di * Yi(1) + (1 - Di) - Y,(0)$ for a random sample of individuals.

- The individual treatment effect, or causal effect, is $Yi(1) - Y1(0)$ but since $Yi(1)$ and $Yj(0)$ are never observed for the same individual.

Theory of causal inference

- an unbiased estimator for the average treatment effect,$E[Y(1) - Yi(O)]$, is available in the difference of the treatment/control averages, $\frac{\sum DiYi}{\sum Di} - \frac{\sum(1-Di)Yi}{\sum(1-Di)}$

- the participation decision is typically modeled by a latent index $D_i^* = \gamma o + Z_i\gamma_1 + vi$ with the observed participation indicator, Di, related to the unobserved latent index, by $Di = \{^{1 \text{ if } D_i^* > 0}_{0 \text{ if } D_i^* \ll 0}$. The response,$Yi$, is related to the treatment via the equation $Y_i = \beta_0 + D_i\beta_1 + \varepsilon_i,$.

- In this notation the counterfactuals are $Y_i\,(0) = \beta_0 + \varepsilon_i$, $Y_i\,(1) = \beta_0 + \beta_1 + \varepsilon_i$,and $Di = 1\,\{\gamma o + Z_i\gamma_1 + vi \ > 0\}$, where $1\,\{.\}$ is the indicator function.

Propensity score matching and Difference in Diffference (PSM-DID)

- Matching pairs the observed outcome of a person in one treatment group with the outcome of the “closest” person in the other treatment group.

- males are paired with males and females are paired with females.

RA: Regression adjustment - know the determinants of the outcome
IPW: Inverse probability weighting - When you know the determinants of treatment status
IPWRA: Inverse probability weighting with regression adjustment -doubly robust estimators
AIPW: Augmented inverse probability weighting - doubly robust estimators
NNM: Nearest-neighbor matching -lots of continuous covariates
PSM: Propensity-score matching - When you know the determinants of treatment status
https://blog.stata.com/2015/08/24/introduction-to-treatment-effects-in-stata-part-2/

- Matching on continuous variables, such as age or weight, can be trickier because of the sparsity of the data. It is unlikely that there are two 45-year-old white males who weigh 193 pounds in a sample. In such situations match subjects who have approximately the same weight and approximately the same age.

1.18.1 Literature on causal inference

- Abadie, A. and Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. econometrica, 74(1), pp.235-267

- Athey, S. and Imbens, G.W., 2019. Machine learning methods that economists should know about. Annual Review of Economics, 11, pp.685-725.

- Imbens, G.W. and Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. Journal of economic literature, 47(1), pp.5-86.

- Imbens, G.W., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. Review of Economics and statistics, 86(1), pp.4-29.

- Imbens, G.W. and Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of econometrics, 142(2), pp.615-635.

- Imbens, G. (2000)). 'The Role of the Propensity Score in Estimating Dose-Response Functions'. Biometrika 87(3): 706–10

- Hirano, K., and G. Imbens (2001) 'Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization'. Health Services and Outcomes Research Methodology, 2(3): 259–78

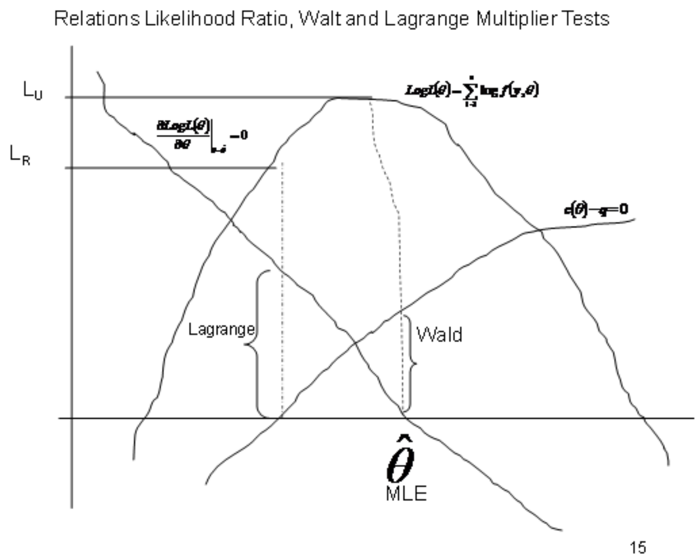
- Imbens, G., and D. Rubin (2015) Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press.

- Unnikrishnan and Imai (2020) Does the old-age pension scheme improve household welfare? Evidence from India. World Development, 134, p.105017

- Abaadie, A. and Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. econometrica, 74(1), pp.235-267.

- Angrist, J. and Imbens, G., 1995. Identification and estimation of local average treatment effects

- Heckman, J., 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. Journal of human resources, pp.441-462.



2 L2: Maximum Likelihood Method

Maximum Likelihood Method (MLE) is more general method of estimation of unknown parameters and applies to both the linear and non-linear models. It also applied to estimate parameters in a system of equations (Wald (1943), Hendry (1971), Rao (1972), Anderson(1974), Phillips (1976,1992), Amemiya (1975, 1977) Lee (1993), Hillier and Armstrong (1999), Aït-Sahalia (2002), Hartley and Mallela (1997), Nielsen (2004), Sweeting (1980), Durham, Gallant, Ait-Sahalia and Brandt (2002), Nielsen (2004), Fukac and Pagan (2010). Main Featurs of a Maximum Likelihood Method include:

- Large sample and more general method
- Appropriate for functions non-linear in parameters
- Unbiased estimates
- Equivalent to the OLS estimates for linear models
- Asymptotically consistent and efficient
- Source of Likelihood Ratio test, Wald Test and LM test

Main Features of a Maximum Likelihood Method
Equivalence of ML to OLS Estimators
Take linear regression model:

$$Y_i = \alpha + \beta X_i + e_i \tag{698}$$

Where the errors are $e_i \sim N\left(0, \sigma^2\right)$

Then the joint distribution or the Likelihood function is given by:

$$L(\alpha,\beta,\sigma) = L(y_1, y_2, ...y_N) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(y_i-\alpha-\beta X_i)^2}{\sigma^2}\right] \tag{699}$$

Take log of this function to get a log-liklihood function.

$$\begin{aligned} LogL(\alpha,\beta,\sigma) &= \sum_{i=1}^{N} -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i-\alpha-\beta X_i)^2}{\sigma^2} \\ &= c - \frac{N}{2}\log\left(\sigma^2\right) - \frac{Q}{2\sigma^2} \end{aligned} \tag{700}$$

Equivalence of ML to OLS Estimators

where $c = \frac{N}{2}\log(2\pi)$ and $Q = \sum_{i=1}^{N}(y_i-\alpha-\beta X_i)^2$

Maximising this likelihood w.r.t. α, β and σ. is equivalent to minimizing Q , which is the negative term in the likelihood function. Therefore the estimators of α, β and σ under the ML method are the same as in the OLS method.

If we substitute the values of $\widehat{\alpha}, \widehat{\beta}$ in the likelihood function, $L(\alpha,\beta,\sigma)$ it just becomes a function of σ

This can be written as

$$\begin{aligned} L\left(\widehat{\alpha},\widehat{\beta},\sigma\right) &= \sum_{i=1}^{N} -\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\frac{\left(y_i-\widehat{\alpha}-\widehat{\beta}X_i\right)^2}{\sigma^2} \\ &= c - N\log\left(\sigma^2\right) - \frac{\widehat{Q}}{2\sigma^2} \end{aligned} \tag{701}$$

Equivalence of ML to OLS Estimators

$$\frac{\partial L\left(\widehat{\alpha},\widehat{\beta},\sigma\right)}{\partial\sigma} = -\frac{N}{\sigma} + \frac{\widehat{Q}}{2\sigma^3} = 0 \Longrightarrow \widehat{\sigma}^2 = \frac{\widehat{Q}}{N} = \frac{RSS}{N} \tag{702}$$

$$\underset{\widehat{\alpha},\widehat{\beta}}{Min}\ Q = \sum \varepsilon_i^2 = \sum\left(Y_i - \widehat{\alpha} - \widehat{\beta}X_{1,i}\right)^2 \tag{703}$$

First order conditions

$$\frac{\partial S}{\partial\widehat{\alpha}} = 0; \frac{\partial S}{\partial\widehat{\beta}} = 0; \tag{704}$$

$$\sum\left(Y_i - \widehat{\alpha}N - \widehat{\beta}X_i\right)(-1) = 0 \tag{705}$$

$$\sum\left(Y_i - \widehat{\alpha}N - \widehat{\beta}X_i\right)(-X_i) = 0 \tag{706}$$

Equivalence of ML to OLS Estimators

$$\sum Y_i = \widehat{\alpha} N + \widehat{\beta} \sum X_i \tag{707}$$

$$\sum Y_i X_i = \widehat{\alpha} \sum X_i + \widehat{\beta} \sum X_i^2 \tag{708}$$

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix} \tag{709}$$

$var(\varepsilon_i) = \widehat{\sigma}^2 = \frac{\sum \widehat{\varepsilon}_i^2}{N-K}$

Compare the corresponding estimate in the OLS .$\widehat{\sigma}^2 = \frac{\widehat{Q}}{N-2} = \frac{RSS}{N-2}$

Thus For large sample OLS and ML estimate are close to each other.

Evaluating the Likelihood Function

Substitituion this value in the likelihood

$$\log L\left(\widehat{\alpha}, \widehat{\beta}, \sigma\right) = c - \frac{N}{2}\log\left(\sigma^2\right) - \frac{Q}{2\sigma^2} = c - \frac{N}{2}\log\left(\frac{\widehat{Q}}{N}\right) - \frac{N}{\widehat{Q}}\frac{Q}{2} \tag{710}$$

$$\log\left(\widehat{\alpha}, \widehat{\beta}, \sigma\right) = c - \frac{N}{2}\log\left(\widehat{Q}\right) - \frac{N}{2} = const - \frac{N}{2}\log\left(\widehat{Q}\right) \tag{711}$$

Take antilog both sides

$$L\left(\widehat{\alpha}, \widehat{\beta}, \sigma\right) = const \times \widehat{Q}^{-\frac{N}{2}} = const \times RSS^{-\frac{N}{2}} \tag{712}$$

2.0.2 Likelihood Ratio Test

This is a general large sample test based on the ML method. Let θ be the set of parameters defining the ML functions, in the above example. (α, β, σ)

$H_0 : \alpha = 0$; and $\beta = 0$ against alternative hypothesis $H_A : \alpha \neq 0$; and $\beta \neq 0$ and or $H_A : \sigma = 0$; and $\sigma \neq 0$ can be tested using the likelihood ratio test. LR test is defined as:

$$\lambda = \frac{\max L(\theta)_R}{\max L(\theta)_{UR}} = \frac{const \times \left(RSS^{-\frac{N}{2}}\right)_R}{\left(const \times RSS^{-\frac{N}{2}}\right)_{UR}} = \left(\frac{RSS_R}{RSS_{UR}}\right)^{-\frac{N}{2}} \tag{713}$$

Where $L(\theta)_R$ is the value of the likelihood function under restriction and $L(\theta)_{UR}$ is the value of likelihood function without restriction. As derived above

Value of λ is likely to be less than one since $L(\theta)_R$ is expected to be less than $L(\theta)_{UR}$.

Taking log both sides and with some rearrangement, this equals

$$-2log_e.\lambda = N.log_e.\left(\frac{RSS_R}{RSS_{UR}}\right) = N\left(log_e RSS_R - log_e RSS_{UR}\right) \tag{714}$$

$2log_e.\lambda$ is distributed χ_k^2 where k is the number of restrictions.

under $H_0 : \alpha = 0$; and $\beta = 0$

$$
\begin{aligned}
-2log_e.\lambda &= N.log_e.\left(\frac{RSS_R}{RSS_{UR}} = \frac{S_{yy}}{S_{yy}\left(1-r^2\right)}\right);\;\; r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2 \sum y_i^2} \\
-2log_e.\lambda &= N.log_e\left(\frac{1}{1-r^2}\right) = -N.log_e\left(1-r^2\right) \qquad (715)
\end{aligned}
$$

This is distributed $.\chi_1^2$

2.0.3 Wald Test

Using notations $S_{yy} = \sum y_i^2$; $S_{xx} = \sum x_i^2$; $S_{xy} = \sum x_i y_i$; $r^2 = \frac{S_{xy}^2}{S_{yy}S_{xx}}$

This is like t-test $t\left(\widehat{\beta}\right) = \frac{\widehat{\beta}-\beta}{SE\left(\widehat{\beta}\right)} = \frac{\widehat{\beta}-\beta}{SE\left(\widehat{\beta}\right)} \sim t_{T-K}$

$$
W = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} \qquad (716)
$$

where $var\left(\widehat{\beta}\right) = \frac{\widehat{\sigma}^2}{S_{xx}}$ This uses RSS of unrestricted Likelihood $\widehat{\sigma}^2 = \frac{S_{yy}\left(1-r^2\right)}{N}$

2.0.4 Lagrange multiplier test

LM test uses RSS from the restricted Likelihood ratio (under $H_0 : \alpha = 0$; and $\beta = 0$)

$$
LM = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2}{\frac{\widehat{\sigma}^2}{S_{xx}}} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2}{\frac{S_{yy}}{N.S_{xx}}} = N.r^2 \qquad (717)
$$

$r^2 = \frac{S_{xy}^2}{S_{yy}S_{xx}}$; ‘This is distributed χ_1^2.

Comparing Wald, LM and LR Tests Lagrange Multiplier Test

LM test uses RSS from the restricted Likelihood ratio (under $H_0 : \alpha = 0$; and $\beta = 0$)

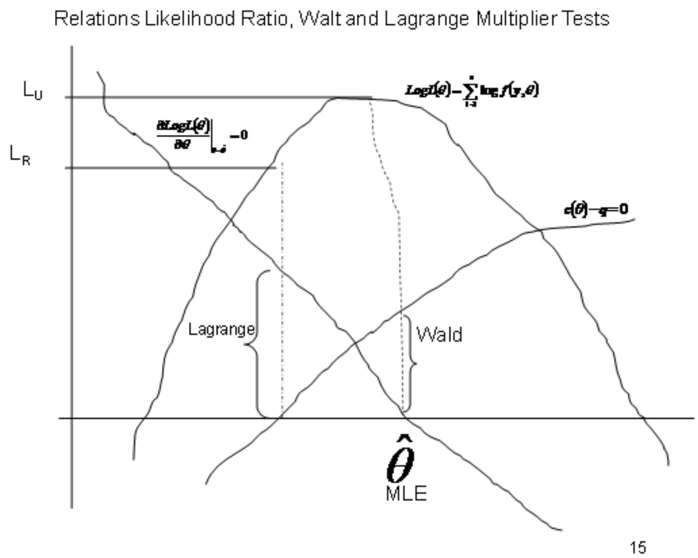
$$
LM = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2}{\frac{\widehat{\sigma}^2}{S_{xx}}} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2}{\frac{S_{yy}}{N.S_{xx}}} = N.r^2 \qquad (718)
$$

$r^2 = \frac{S_{xy}^2}{S_{yy}S_{xx}}$

This is distributed χ_1^2.

Wald uses RSS of unrestricted Likelihood $\widehat{\sigma}^2 = \frac{S_{yy}\left(1-r^2\right)}{N}$

$$
W = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2}{\frac{S_{yy}(1-r^2)}{N.S_{xx}}} = \frac{N.r^2}{(1-r^2)} \qquad (719)
$$



This is distributed χ_1^2
Liklihood ratio test (Neyman-Pearson (1928)): $-2log_e.\lambda = N.log_e\left(\frac{1}{1-r^2}\right) = -N.log_e\left(1-r^2\right)$
Wald Test (1943): $W = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} = \frac{N.r^2}{(1-r^2)}$
LM test Rao (1948)): $LM = \frac{\left(\widehat{\beta}\right)^2}{var\left(\widehat{\beta}\right)} = N.r^2$

$$W \geqslant LR \geqslant LM \tag{720}$$

$$\begin{aligned} \frac{W}{N} &= \frac{.r^2}{(1-r^2)} \geqslant \frac{LR}{N} = log_e\left(\frac{1}{1-r^2}\right) \\ &= log_e\left(1+\frac{W}{N}\right) \geqslant \frac{LM}{N} = r^2 = \frac{W}{N} / \left(1+\frac{W}{N}\right) \end{aligned} \tag{721}$$

with $\frac{W}{N} = x$ it fulfills the inequality $\quad x \geqslant log_e\left(1+x\right) \geqslant \frac{x}{(1+x)}$

2.0.5 Newton Ralphson Algorithm

Consider observations $(y_1, y_2, ...y_N)$ with a density function $f\left(y,\theta\right)$ and $L\left(\theta\right)$ The log likelihood is given by

$$LogL\left(\theta\right) = \sum_{i=1}^{N} \log f\left(y,\theta\right) \tag{722}$$

$$\frac{\partial LogL\left(\theta\right)}{\partial\theta}|_{\theta=\widehat{\theta}} = 0 \tag{723}$$

Iteration procedure involves θ updating in successive iteration starting from θ_0 as $\delta\theta = (\theta - \theta_0)$

Newton Ralphson Algorithm

$$\frac{\partial LogL\,(\theta)}{\partial\theta} = \frac{\partial LogL\,(\theta)}{\partial\theta_0} + \delta\theta\frac{\partial^2 LogL\,(\theta)}{\partial\theta_0^2} = 0 \tag{724}$$

$$(\theta - \theta_0) = \delta\theta = \left[\frac{\partial^2 LogL\,(\theta)}{\partial\theta_0^2}\right]^{-1}\left(\frac{\partial LogL\,(\theta)}{\partial\theta_0}\right) \tag{725}$$

Then continue the correction process until . $\theta_n \longrightarrow \theta_{n+1}$

Gauss-Newton Algorithm

- Above algorithm is close to the Gauss-Newton algorithm used frequently to estimate the parameters of a nonlinear equation $Y = f(X_1, X_2, ...Xk; .\beta_1, ...\beta_{p,0})$ with the Taylor series approximations explained by Pindyck and Robinfeld (1998) as:

$Y = f(X_1, X_2, ...Xk; .\beta_{1,0}......\beta_{p,0}) + \sum\limits_{i=1}^{p}\left(\frac{df}{\partial\beta_i}\right)_0(\beta_i - \beta_0)$
$+\frac{1}{2}\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{p}\left(\frac{d^2 f}{\partial\beta_i\partial\beta_j}\right)_0\left(\beta_i - \beta_{i,0}\right)\left(\beta_j - \beta_{j,0}\right) + ... + \varepsilon.$

- Linear approximation of this only involves the first term $Y - f(X_1, X_2, ...Xk; .\beta_{1,0}......\beta_{p,0}) + \sum\limits_{i=1}^{p}\left(\frac{df}{\partial\beta_i}\right)_0(\beta_i - \beta_0) = \sum\limits_{i=1}^{p}\left(\frac{df}{\partial\beta_i}\right)_0(\beta_i - \beta_0) + ... + \varepsilon..$

- The ML is performed iteratively until the parameters converge; $\left|\frac{\beta_{i,j+1}-\beta_{i,j}}{\beta_{i,j}}\right| < \delta$ for any small value of δ. Alternatively the convergence is checked using a damping factor $\alpha, \beta_{i,j+1} = \beta_{i,j} + \alpha\left(\widehat{\beta}_{i,j+1} - \beta_{i,j}\right)$.

Parametarised Expectation Algorithm

- Computations of the stochastic intertemporal optimal accumulation model follows non-linear iterative algorithm very popular in econometrics and growth models (Han and Marcet (1988)).

- The optimality condition in the stochastic asset accumulation model here are governed by the first order Euler equation $\frac{\beta C_{t,p}^{\kappa}(1+r(1-t_k(p)))}{C_{t+1,p}^{\kappa}} = 1$ relates $C_{t,p}$ and $C_{t+1,p}$ and explains the process of saving and asset accumulation.

- It essentially gives the marginal rate of substitution between the current and the future consumptions for all states and periods. In the presence of transfers, capital and labour income taxes $\frac{1}{C_{t+1,p}^{\kappa}}$ term is estimated by following parameterised expectation function: $\psi_{t,p} = \exp(\delta_{0,t} + \delta_{1,t}\ln\left(x_1 z_1\right)\theta_{t,p} + \delta_{2,t}\ln\left(x_2 z_2\right)\in_{t,p} + \delta_{3,t}\ln\left(x_3 z_3\right)k_{t,p}$

Parametarised Expectation Algorithm

- $\psi_{t,p} = \exp(\delta_{0,t} + \delta_{1,t}\ln\left(x_1 z_1\right)\theta_{t,p} + \delta_{2,t}\ln\left(x_2 z_2\right)\in_{t,p} + \delta_{3,t}\ln\left(x_3 z_3\right)k_{t,p}$

- $+\delta_{4,t}\ln\left(x_4z_4\right)TR_{t,p}+\delta_{5,t}\ln\left(x_5z_5\right)TK_{t,p}+\delta_{6,t}\ln\left(x_6z_6\right)TL_{t,p}$. Now the $\frac{1}{C^{\kappa}_{t+1,p}}$is substituted by $\psi_{t,p}$ as $\beta C^{\kappa}_{t,p}\left(1+r\left(1-t_k(p)\right)\right)\psi_{t,p}=1$. The model converges to true solutions when the values of $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ are equal up to an arbitrarily small number.
- For quadratic approximation a regression is estimated to minimise the sum-squared errors between $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ as $SS=\sum_{t=1}^{T}\sum_{p=1}^{P}\left(\frac{1}{C^{\kappa}_{t+1,p}}-\psi_{t,p}\right)^2$ and the iteration continues until this sum of error square is very small.

Parametarised Expectation Algorithm

- Minor iterations are run to evaluate the expectation $\psi_{t,p}$ and the original Euler equation.
- Then the major iterations are run until the difference between $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ is statistically insignificant.
- Average values and standard deviations of earnings, wealth, income, consumption and savings ,$e_{t,p}$ $W_{t,p},W_{t,p},Y_{t,p},C_{t,p}$,and $S_{t,p}$ are computed when the approximation is close.
- It is possible to fit the confidence intervals and do all other statistical tests with the mean and variances computed in this manner. In a nutshell this algorithm involves following steps

Parametarised Expectation Algorithm
In a nutshell this algorithm involves following steps

1. Generate a series of stochastic earning process $\theta_{t,p}$.
2. choose initial $\delta_{i,t}$ for $\psi_{t,p}$ function; replace $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ in the Euler equation to generate series of consumption $C_{t,p}$.
3. compute the difference of the values of $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ and fit a quadratic error sum minimisation routines until $\frac{1}{C^{\kappa}_{t+1,p}}$ and $\psi_{t,p}$ converge.
4. compute the mean, standard deviations and confidence intervals for model variables ,$e_{t,p}$ $W_{t,p},W_{t,p},Y_{t,p},C_{t,p}$,and $S_{t,p}$.
5. compare results for different tax experiments.

2.0.6 MLE: example1

An urn contains N balls and N1 of them are red. The Bernauli probability distribution for red balls in a draw as a discrete likelihood function is given by

$$p(n)=p^{N_1}(1-p)^{N-N_1} \tag{726}$$

The log likelihood function for this is given by:

$$ln(p)=N_1\ln(p)+(N-N_1)\ln\left(1-p\right) \tag{727}$$

First order conditions for maximisation

$$\frac{\partial \ln(p)}{\partial p} = \frac{N_1}{p} - \frac{(N - N_1)}{(1 - p)} = 0 \tag{728}$$

This implies $\frac{N_1}{p} - \frac{(N-N_1)}{(1-p)} = 0$;or $(1 - p)N_1 = p(N - N_1)$and

$$p = \frac{N_1}{N} \tag{729}$$

MLE: example1

- Second order condition for maximisation

$$\frac{\partial^2 \ln(p)}{\partial p^2} = -\frac{N_1}{p^2} - \frac{(N - N_1)}{(1 - p)^2} < 0 \tag{730}$$

- It proves that the estimate $p = \frac{N_1}{N}$ maximises the likelihood function.
- Thus for a given distribution of data, the maximum likelihood estimation procedure involves determining the unknown parameters (here p) that maximises the likelihood of the observed data.

2.0.7 MLE: Example 2

- Take a Poissoin likelihood denstity fucntion for $f\left(y_1, y_2, ..., y_N/\theta\right) = \prod_{i=1}^{N} f(y/\theta)$; poisson density $f\left(y/\theta\right) = \frac{e^{-n\theta}}{y_1!}$

MLE: example2

$$L(\theta/y) = \frac{e^{-n\ \theta}\theta^{\sum\limits_{i=1} y_i}}{\prod\limits_{i=1}^{N} y_1!} \tag{731}$$

Log liklihood of this is given by

$$\ln L(\theta/y) = -n\theta + \ln\theta \sum y_i - \sum \ln\left(y_i!\right) \tag{732}$$

data example from Greene (485): $n = 10$; y: $(5, 0, 1, 1, 0, 3, 2, 3, 4, 1)$ $\sum y_i = 20$; $\sum y_i! = 207,360$

$$\ln L(\theta/y) = -10\theta + 20\ln\theta - 12.242 \tag{733}$$

From the first order condition θ that maximises this function is given by $\frac{\partial \ln L(\theta/y)}{\partial\theta} = -10 + \frac{20}{\theta} = 0$ $\theta = \frac{20}{10} = 2$

This is maximum is ascertained by the negative second derivative: $\frac{\partial^2 \ln L(\theta/y)}{\partial\theta^2} = -\frac{20}{\theta^2} < 0$
MLE: example3

Let Y_T be a random sample normally distributed with parameters $N\left(\beta,\sigma^2\right)$. Here parameter $\theta=\left(\beta,\sigma^2\right)$. Log likelihood function is defined as

$$lnL(\theta/y)=\ln\left\{\prod_{i=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(y_i-\beta)^2}{\sigma^2}\right]\right\}=\ln\left\{(2\pi)^{-\frac{T}{2}}\left(\sigma^2\right)^{-\frac{T}{2}}\exp\left[-\frac{1}{2}\sum_{i=1}^{T}\frac{(y_i-\beta)^2}{\sigma^2}\right]\right\}$$

$$lnL(\theta/y)=\left\{-\frac{T}{2}\ln(2\pi)-\frac{T}{2}\ln\left(\sigma^2\right)-\frac{1}{2}\sum_{i=1}^{T}\frac{(y_i-\beta)^2}{\sigma^2}\right\}\tag{734}$$

The first order conditions for optimisations are :

$$\frac{\partial lnL(\theta/y)}{\partial\beta}=\frac{1}{\sigma^2}\left(\sum_{i=1}^{T}y_i-T\beta\right)=0\tag{735}$$

$$\frac{\partial lnL(\theta/y)}{\partial\sigma^2}=-\frac{T}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{T}\left(y_i-\beta\right)^2=0\tag{736}$$

MLE: example3

The first order conditions imply

$$\sum_{i=1}^{T}y_i=T\beta;\quad \beta=\frac{\sum_{i=1}^{T}y_i}{T}=\overline{y}\quad\text{and}\quad\sigma^2=\frac{\sum_{i=1}^{T}\left(y_i-\beta\right)^2}{T}\tag{737}$$

and the second order conditions are

$$\frac{\partial^2 lnL(\theta/y)}{\partial\beta^2}=-\frac{T}{\sigma^2}<0\tag{738}$$

$$\frac{\partial^2 lnL(\theta/y)}{\partial\left(\sigma^2\right)^2}=-\frac{T}{2\sigma^4}-\frac{1}{2\sigma^6}\sum_{i=1}^{T}\left(y_i-\beta\right)^2<0\tag{739}$$

$$\frac{\partial lnL(\theta/y)}{\partial\sigma^2\partial\beta}=-\frac{1}{2\sigma^2}\sum_{i=1}^{T}\left(y_i-T\beta\right)<0\tag{740}$$

The second order conditions for maximisation are

$$\begin{bmatrix}\frac{\partial^2 lnL(\theta/y)}{\partial\beta^2} & \frac{\partial lnL(\theta/y)}{\partial\sigma^2\partial\beta}\\ \frac{\partial lnL(\theta/y)}{\partial\beta\partial\sigma^2} & \frac{\partial^2 lnL(\theta/y)}{\partial\left(\sigma^2\right)^2}\end{bmatrix}=\begin{bmatrix}-\frac{T}{\sigma^2} & 0\\ 0 & -\frac{T}{2\sigma^2}\end{bmatrix}\tag{741}$$

This is a negative definite matrix, Hessians have alternate signs, that fulfills the conditions for maximisation.

Likelihood ratio test $\lambda=\frac{L_R}{L_U}\sim\chi^2_{df=\#\ r}$

Lagrange test is obtained from restricted maximum likelihood: $\ln L^*\left(\theta\right)=\ln L\left(\theta\right)+\lambda\left[c(\theta)-q\right]$

Wald test needs only unrestricted estimation:

$W=\left(c(\widehat{\theta})-q\right)'\left[Assmptvar\left(c(\widehat{\theta})-q\right)\right]^{-1}\left(c(\widehat{\theta})-q\right)'\sim\chi^2_r$

2.0.8 MLE: Example 3

(Adapted from Greene (496)):

Let productivity (pleasure), as a function of income (y) and education (x) be given by the following exponfunction:

$$f\left(y_i, x_i, \beta\right) = \frac{1}{\beta + x_i} \exp - \frac{y_i}{\beta + x_i} \tag{742}$$

This is a non-linear function. Liklihood for such dgp is given by L

$$L(\theta) = \prod_{i=1}^{N} f\left(y_i, x_i, \beta\right) = \prod_{i=1}^{N} \left[\frac{1}{\beta + x_i} \exp - \frac{y_i}{\beta + x_i}\right] \tag{743}$$

Log likelihood of this :

$$lnL(\theta) = -\sum_{i=1}^{N} \ln\left(\beta + x_i\right) - \sum_{i=1}^{N} \frac{y_i}{\beta + x_i} \tag{744}$$

$$\frac{\partial lnL(\theta)}{\partial \beta} = -\sum_{i=1}^{N} \frac{1}{\left(\beta + x_i\right)} + \sum_{i=1}^{N} \frac{y_i}{\left(\beta + x_i\right)^2} = 0 \tag{745}$$

The assymptotic variance of the MLE is given by the second derivative (information matrix) $\frac{\partial^2 lnL(\theta)}{\partial \beta^2} = \sum_{i=1}^{N} \frac{1}{\left(\beta + x_i\right)^2} - 2\sum_{i=1}^{N} \frac{y_i}{\left(\beta + x_i\right)^3}$

BHHH(Berndt, Hall, Hall and Hauseman (1974) estimate BHHH(Berndt, Hall, Hall and Hauseman (1974) estimate $BHHH = \frac{1}{\sum_{i=1}^{N}\left[\frac{1}{\left(\beta + x_i\right)} + \frac{y_i}{\left(\beta + x_i\right)^2}\right]^2}$

3 BHHH Algorithm

1. The Berndt–Hall–Hall–Hausman (BHHH) algorithm is a numerical optimization algorithm similar to the Newton–Raphson algorithm, but it replaces the observed negative Hessian matrix with the outer product of the gradient.

2. It is named after the four originators: Ernst R. Berndt, Bronwyn Hall, Robert Hall, and Jerry Hausman.

3. If a nonlinear model is fitted to the data one often needs to estimate coefficients through optimization.

- A number of optimisation algorithms have the following general structure. Suppose that the function to be optimized is $Q(\beta)$.

- Then the algorithms are iterative, defining a sequence of approximations, β_k given by

$$\beta_{k+1} = \beta_k - \lambda_k A_k \frac{\partial Q}{\partial \beta}(\beta_k)$$

where β_k is the parameter estimate at step k, and λ_k is a parameter (called step size) which partly determines the particular algorithm.

- For the BHHH algorithm λ_k is determined by calculations within a given iterative step, involving a line-search until a point β_{k+1} is found satisfying certain criteria. In addition, for the BHHH algorithm, Q has the form $Q = \sum_{i=1}^{n} Q_i$ and A is calculated using

$$A_k = \left[\sum_{i=1}^{n} \frac{\partial Q}{\partial \beta}(\beta_k) \frac{\partial Q}{\partial \beta}(\beta_k)'\right]^{-1}$$

In other cases, e.g. Newton–Raphson, A_k can have other forms. The BHHH algorithm has the advantage that, if certain conditions apply, convergence of the iterative procedure is guaranteed.

3.0.9 DFP Algorithm

The Davidon–Fletcher–Powell formula (or DFP; named after William C. Davidon, Roger Fletcher, and Michael J. D. Powell) finds the solution to the secant equation that is closest to the current estimate and satisfies the curvature condition. It was the first quasi-Newton method to generalize the secant method to a multidimensional problem. This update maintains the symmetry and positive definiteness of the Hessian matrix.

Given a function $f(x)$ its gradient ∇f, and positive-definite Hessian matrix B, the Taylor series is $f(x_k + s_k) = f(x_k) + \nabla f(x_k)^T s_k + \frac{1}{2} B s_k + \ldots.,$

and the Taylor series of the gradient itself (secant equation)

$\nabla f(x_k + s_k) = \nabla f(x_k)^T s_k + B s_k \ldots$

is used to update B.

The DFP formula finds a solution that is symmetric, positive-definite and closest to the current approximate value of B_k:

$B_{k+1} = \left(I - \gamma_k y_k s_k^T\right) B_k \left(I - \gamma_k s_k y_k^T\right) + \gamma_k y_k y_k^T$

where $y_k = \nabla f(x_k + s_k) - \nabla f(x_k), \quad \gamma_k = \frac{1}{y_k^T s_k}$

and B_k is a symmetric and positive-definite matrix.

The corresponding update to the inverse Hessian approximation

$H_k = B_k^{-1}$ is given by

$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}$

B is assumed to be positive-definite, and the vectors s_k^T and y must satisfy the curvature condition

$s_k^T y_k = s_k^T B s_k > 0$

The DFP formula is quite effective, but it was soon superseded by the BFGS formula, which is its dual (interchanging the roles of y and s).

3.0.10 BFGS algorithm

BFGS algorithm-1

The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is an iterative method for solving unconstrained nonlinear optimization problems.

The BFGS method belongs to quasi-Newton methods, a class of hill-climbing optimization techniques that seek a stationary point of a (preferably twice continuously differentiable) function. For such problems, a necessary condition for optimality is that the gradient be zero. Newton's method and the BFGS methods are not guaranteed to converge unless the function has a quadratic Taylor expansion near an optimum. However, BFGS can have acceptable performance even for non-smooth optimization instances.

In quasi-Newton methods, the Hessian matrix of second derivatives is not computed. Instead, the Hessian matrix is approximated using updates specified by gradient evaluations (or approximate gradient evaluations). Quasi-Newton methods are generalizations of the secant method to find the root of the first derivative for multidimensional problems. In multi-dimensional problems, the secant equation does not specify a unique solution, and quasi-Newton methods differ in how they constrain the solution. The BFGS method is one of the most popular members of this class. Also in common use is L-BFGS, which is a limited-memory version of BFGS that is particularly suited to problems with very large numbers of variables (e.g., >1000). The BFGS-B variant handles simple box constraints.

The algorithm is named after Charles George Broyden, Roger Fletcher, Donald Goldfarb and David Shanno.

The optimization problem is to minimize $f(x)$, where x is a vector in $\mathbb{R}^n$, and f is a differentiable scalar function. There are no constraints on the values that x can take.

The algorithm begins at an initial estimate for the optimal value x_0 and proceeds iteratively to get a better estimate at each stage.

The search direction p_k at stage k is given by the solution of the analogue of the Newton equation:

$B_k p_k = -\nabla f(x_k)$

where B_k is an approximation to the Hessian matrix, which is updated iteratively at each stage, and $\nabla f(x_k)$ is the gradient of the function evaluated at x_k. A line search in the direction p_k is then used to find the next point x_{k+1} by minimizing $f(x_k + \alpha p_k)$ over the scalar $\alpha > 0$.

The quasi-Newton condition imposed on the update of B_k is

$B_{k+1}\left(x_{k+1} - x_k\right) = \nabla f(x_{k+1}) - \nabla f(x_k)$

Let $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and $s_k = x_{k+1} - x_k$, then B_{k+1} satisfies $B_{k+1} s_{k+1} = y_k$, which is the secant equation. The curvature condition $s_k^T y_k > 0$ should be satisfied for B_{k+1} to be positive definite, which can be verified by pre-multiplying the secant equation with s_k^T. If the function is not strongly convex, then the condition has to be enforced explicitly.

Instead of requiring the full Hessian matrix at the point x_{k+1} to be computed as B_{k+1}, the approximate Hessian at stage k is updated by the addition of two matrices:

$B_{k+1} = B_k + U_k + V_k$

Both U_k and V_k are symmetric rank-one matrices, but their sum is a rank-two update matrix. BFGS and DFP updating matrix both differ from its predecessor by a rank-two matrix. Another simpler rank-one method is known as symmetric rank-one method, which does not guarantee the positive definiteness. In order to maintain the symmetry and positive definiteness of B_{k+1}, the update form can be chosen as

$B_{k+1} = B_k + \alpha\, \mathrm{u}\, \mathrm{u}^T + \beta v v^T$. Imposing the secant condition, $B_{k+1}s_{k+1} = y_k$. Choosing $u = y_k$ and $v = B_{k+1}s_{k+1}$, we can obtain:

$\alpha = \frac{1}{y_k^T s_k}$ and $\beta = -\frac{1}{s_k^T B_k s_k}$

Finally, we substitute α and β into $B_{k+1} = B_k + \alpha\, \mathrm{u}\, \mathrm{u}^T + \beta v v^T$ and get the update equation of B_{k+1}:

$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k}$

From an initial guess x_0 and an approximate Hessian matrix B_0 the following steps are repeated as x_k converges to the solution:

Obtain a direction p_k by solving $B_k p_k = -\nabla f(x_k)$.

Perform a one-dimensional optimization (line search) to find an acceptable stepsize α_k in the direction found in the first step, so $\alpha_k = \arg\min f(x_k + \alpha p_k)$

Set $s_k = \alpha p_k$ and update $x_{k+1} = x_k + s_k$

$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k}$

$f(x)$ denotes the objective function to be minimized. Convergence can be checked by observing the norm of the gradient,$\|\nabla f(x)\|$. If B_0 is initialized with $B_0 = I$, the first step will be equivalent to a gradient descent, but further steps are more and more refined by B_k, the approximation to the Hessian.

The first step of the algorithm is carried out using the inverse of the matrix B_k, which can be obtained efficiently by applying the Sherman–Morrison formula to the step 5 of the algorithm, giving

$B_{k+1}^{-1} = \left(I - \frac{\left(s_k y_k^T\right)}{y_k^T s_k}\right) B_k^{-1} \left(I - \frac{\left(y_k s_k^T\right)}{y_k^T s_k}\right) + \frac{\left(s_k s_k^T\right)}{y_k^T x_k}$

This can be computed efficiently without temporary matrices, recognizing that B_k^{-1} is symmetric, and that $y_k^T B_k^{-1} y_k$ and $s_k^T y_k$ are scalars, using an expansion such as

$B_{k+1}^{-1} = B_k^{-1} + \frac{\left(s_k^T y_k + y_k^T B_k^{-1} y_k\right)\left(s_k s_k^T\right)}{\left(s_k^T y_k\right)^2} - \frac{B_k^{-1} y_k s_k^T + s_k^T y_k^T B_k^{-1}}{y_k^T s_k}$

In statistical estimation problems (such as maximum likelihood or Bayesian inference), credible intervals or confidence intervals for the solution can be estimated from the inverse of the final Hessian matrix. However, these quantities are technically defined by the true Hessian matrix, and the BFGS approximation may not converge to the true Hessian matrix.

Exercises:

1. Construct STATA program file (*.do file) to implement a BHHH for a cross section analyis; DFP and BFGS algorithms for time series or panel data analysis.

2. Construct Eviews program file (*.do file) to implement a BHHH for a cross section analyis; DFP and BFGS algorithms for time series or panel data analysis.

3. Construct RATS program file (*.do file) to implement a BHHH for a cross section analyis; DFP and BFGS algorithms for time series or panel data analysis.

4. Construct R program file (*.do file) to implement a BHHH for a cross section analyis; DFP and BFGS algorithms for time series or panel data analysis.

5. Review three articles from Econometrics or applied econometric journals that apply BHHH, DFP and BFGS algorithms

MLE: AR(1) Likelihood function for $AR(1)$ process

$$Y_t = c + \phi Y_{t-1} + \epsilon_t \tag{746}$$

Its mean and variances are

$$\mu = \frac{c}{1-\phi}; \quad \sigma_y^2 = \frac{\sigma^2}{1-\phi^2} \tag{747}$$

Likelihood functions:
The disnsity of the first orbservation:

$$\begin{aligned} L_{y_1}(y_1,\theta) &= \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_1-\mu)^2}{\sigma_y^2}\right] \\ &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\phi^2}}} \exp\left[-\frac{1}{2}\frac{\left(y_1-\frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}}\right] \end{aligned} \tag{748}$$

density for the second observation given that the first has been observed:

$$L_{y_2/y_1}(y_2/y_1,\theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2-c-\phi y_1)^2}{\sigma_y^2}\right] \tag{749}$$

$$L(y_t/y_{t-1},y_{t-2},....y_{1,}\theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2-c-\phi y_{t-1})^2}{\sigma_y^2}\right] \tag{750}$$

Likelihood of the complete function

$$\begin{aligned} & L_{y_t,y_{t-1},y_{t-2},....y_1}(y_t/y_{t-1},y_{t-2},....y_{1,}\theta) \\ = \; & L_{y_1}(y_1,\theta).\prod_{t=2}^{T} L(y_t/y_{t-1},y_{t-2},....y_{1,}\theta) \end{aligned} \tag{751}$$

Logliklihood function

$$\Im(\theta) = \log L_{y_1}(y_1,\theta) + \sum_{t=2}^{T} \log L(y_t/y_{t-1},y_{t-2},....y_{1,}\theta) \tag{752}$$

Sustiting the relevant functions, the likelihood for sample T Gaussian process is

$$\begin{aligned} \Im(\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right)\left[-\frac{1}{2}\frac{\left(y_1-\frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}}\right] + \\ & -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2} - \sum_{t=2}^{T}\left[\frac{(y_t-c-\phi y_{t-1})^2}{2\frac{\sigma^2}{1-\phi^2}}\right] \end{aligned} \tag{753}$$

MLE: MA(1) $MA(1)$ process

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1} \tag{754}$$

$$L_{y_t}(y_{t/\epsilon_{t-1}}, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_1 - \mu - \theta\epsilon_{t-1})^2}{\sigma_y^2}\right] \tag{755}$$

$$\epsilon_1 = y_1 - \mu \tag{756}$$

$$L_{y_2/y_1}(y_2/y_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2 - c - \phi y_1)^2}{\sigma_y^2}\right] \tag{757}$$

$$\epsilon_2 = y_2 - \mu - \theta\epsilon_1 \tag{758}$$

$$\epsilon_2 = y_2 - \mu - \theta\epsilon_1 \tag{759}$$

$$\epsilon_t = y_t - \mu - \theta\epsilon_{t-1} \tag{760}$$

$$L(y_t/y_{t-1}, y_{t-2}, ...y_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_y^2}\right] \tag{761}$$

Conditional log-likelihood for MA(1)

$$\Im(\theta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=2}^{T}\left[\frac{1}{2}\frac{\epsilon_t^2}{\frac{\sigma^2}{1-\phi^2}}\right] \tag{762}$$

MLE: MA(p,q) Similarly for the MA(p,q)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + ... + \theta_p\varepsilon_{t-q} \tag{763}$$

$$\Im(\theta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=2}^{T}\left[\frac{1}{2}\frac{\epsilon_t^2}{\frac{\sigma^2}{1-\phi^2}}\right] \tag{764}$$

MLE: ARCH Joint realisations of e_1 , e_2, e_T is given by

$$L = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp\left(\frac{-e_t^2}{\sqrt{2\sigma^2}} \right) \tag{765}$$

in log form

$$\ln L = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(e_t^2\right) \tag{766}$$

where $e_t = Y_t - \beta_0 - \beta_1 X_t$.

$$\ln L = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln\left(\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(Y_t - \beta_0 - \beta_1 X_t\right)^2 \tag{767}$$

First order conditions for maximisations are

$$\frac{\ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{T}{2\sigma^4} \sum_{t=1}^{T} \left(Y_t - \beta_0 - \beta_1 X_t\right)^2 \tag{768}$$

$$\frac{\ln L}{\partial \beta_1} = \frac{T}{2\sigma^2} \sum_{t=1}^{T} \left(Y_t X_t - \beta_0 X_t - \beta_1 X_t^2\right) \tag{769}$$

For a model without intercept this gives

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T} \left(e_t^2\right)}{T}; \quad \widehat{\beta}_1 = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} \tag{770}$$

This maximum likelyhood method applied to the ARCH errors generates

$$lnL = -n\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \ln\left(\alpha_0 + \alpha_1 e_{t-1}^2\right) - \frac{1}{2} \sum_{t=1}^{T} \frac{e_t^2}{\alpha_0 + \alpha_1 e_{t-1}^2} \tag{771}$$

Algorithm of ARCH process is non-linear iterative procedure. It is not possible to estimate β_0, β_1 n $Y_t = \beta_0 + \beta_1 X_{1,t} + e_t$ without knowing e_t here errors are not normal. mean of e_t can still be zero but its variance is modelled in the variance equation.

Thus the estimation is higly non-liner and maximul likelihoold method is used to estimate this.

First start with the initial values of $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$, and e_t .. e_{t-S} . Estimate $h_t = \sigma_t^2$.

Secondly, estimate $\widehat{\beta}_0$, $\widehat{\beta}_1$ based on $\widehat{\sigma}_t^2$. Then estimate $\widehat{e}_t$ from this estimate the variance $\widehat{\sigma}_t^2$ and new vlaues of $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$.

Then continue the process until the values of $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$ converge.

MLE: GARCH

It is possible to add other explanatory variables in the variance equation

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \sum_{j=1}^{p} \delta^j h_t + \sum_{j=k}^{m} \mu_k X_k \tag{772}$$

All forms of GARCH modesl are estimated using iterative Maximum Likelihood procedure.

$$lnL\left(\theta\right) = \frac{1}{T}\sum_{t=1}^{T} l_t = -\frac{1}{2T}\sum_{t=1}^{T} \ln\left(h_t\right) - \frac{1}{2T}\sum_{t=1}^{T}\frac{e_t^2}{h_t} \tag{773}$$

Bayesian Likelihood

$$L(\beta, \sigma^2) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp - \frac{1}{2\sigma^2}\left[\sum_{i}^{N}(Y - \widehat{\beta}X)^2 - \left(\beta - \widehat{\beta}\right)^2 \sum_{i}^{N} X_i^2\right] \tag{774}$$

$s = \frac{\sum (Y - \widehat{\beta}X)^2}{N-1} = vs^2$ with $v = N - 1$

$$\sum_{i}^{N}\left(Y - \beta X\right)^2 = -vs^2 + \left(\beta - \widehat{\beta}\right)^2 \sum_{i}^{N} X_i^2 \tag{775}$$

$$L(\beta, \sigma^2) = \frac{1}{\left(2\pi\right)^{\frac{N}{2}}} \left\{ \frac{1}{\sqrt{\sigma^2}} \exp\left[-\frac{1}{2} \frac{\left(\beta - \widehat{\beta}\right)^2}{\sigma^2 \left(\sum_{i}^{N} X_i^2\right)^{-1}} \right] \frac{1}{\left(\sqrt{\sigma^2}\right)^v} \exp\left[-\frac{vs^2}{2\sigma^2}\right] \right\} \tag{776}$$

MLE: Limited Information Maximum Likelihood Single equation method for the simultaneous equation method (like 2SLS); For instance, example in Maddala(2001)

$$y_1 = b_1 y_2 + c_1 z_1 + c_2 z_2 + u_1 \tag{777}$$

$$y_2 = b_2 y_1 + c_3 z_3 + u_2 \tag{778}$$

define

$$y_1^* = y_1 - b_1 y_2 = c_1 z_1 + c_2 z_2 + u_1 \tag{779}$$

First regress y_1^* on z_1, z_2 .; get the RSS1
Then regress y_1^* on $z_1, z_2 .z_3$ get RSS2
Estimate b_1 to minimise $\frac{RSS1}{RSS2}$
Determine c_1, c_2 after estimating b_1 . LIML and 2SLS are almost same for the exactly identified syste.

3.1 Full Information Maximum Likelihood Method (FIML)

Classic example developed by Klein (1956) and reiterated by Koutsoyiannis (1973) is the easiest way to understand the FIML

$$C = a_0 + a_1(Y - T) + a_2 Y_{t-1} + u_1 \tag{780}$$

$$I = b_0 + b_1 Y + b_2 K_{t-1} + b_3 r + b_4 E + u_2 \tag{781}$$

$$M = c_0 + c_1 Y + c_2 P_{t-1} + u_3 \tag{782}$$

$$Y = C + I + G + E - M \tag{783}$$

substituting out the income equation

$$C = \frac{a_0}{1-a_1} + \frac{a_1}{1-a_1}(I + G + E - M - T) + \frac{a_2}{1-a_1} Y_{t-1} + \frac{u_1}{1-a_1} \tag{784}$$

Full Information Maximum Likelihood Method (FIML)

$$\begin{aligned} I \quad = \quad & \frac{b_0}{1-b_1} + \frac{b_1}{1-b_1}(C + G - M) + \frac{b_2}{1-b_1} K_{t-1} \\ & + \frac{b_3}{1-b_1} r + \frac{b_1 + b_4}{1-b_1} E + \frac{u_2}{1-b_1} \end{aligned} \tag{785}$$

$$M = \frac{c_0}{1+c_1} + \frac{c_1}{1+c_1}(C + I + G + E) + \frac{c_2}{1+c_1} P_{t-1} + \frac{u_3}{1+c_1} \tag{786}$$

define the reduced form parameters as $\frac{a_0}{1-a_1} = \alpha_0$; $\frac{a_1}{1-a_1} = \alpha_1$; $\frac{a_2}{1-a_1} = \alpha_3$; $\frac{u_1}{1-a_1} = u_1^*$;
$\frac{b_0}{1-b_1} = \beta_0$; $\frac{b_1}{1-b_1} = \beta_1$; $\frac{b_2}{1-b_1} = \beta_2$; $\frac{b_3}{1-b_1} = \beta_3$; $\frac{b_1+b_4}{1-b_1} = \beta_4$; $\frac{u_2}{1-b_1} = u_2^*$
$\frac{c_0}{1+c_1} = \gamma_0$; $\frac{c_1}{1+c_1} = \gamma_1$; $\frac{c_2}{1+c_1} = \gamma_2$; $\frac{u_3}{1+c_1} = u_3^*$

Full Information Maximum Likelihood Method (FIML)

Simplifying the notations for the endogenous and exogenous variables

Endogeneous variables: $y_1 = C$; $y_2 = I$; $y_3 = M$;

Exogenous variables: $z_1 = G; z_2 = E; z_3 = K_{t-1}; z_4 = T; z_5 = Y_{t-1}; z_6 = r; z_7 = P_{t-1}$;

$y_1 = \alpha_0 + \alpha_1(y_2 + z_1 + z_2 - y_3 - z_4) + \alpha_2 z_5 + u_1^*$; $y_2 = \beta_0 + \beta_1(y_1 + z_1 - y_3) + \beta_2 z_3 + \beta_3 z_6 + \beta_4 z_2 + u_2^*$ and

$y_3 = \gamma_0 + \gamma_1(y_1 + y_2 + z_1 + z_2) + \gamma_2 z_7 + u_3^*$

$$y_1 = \alpha_0 + \alpha_1(y_2) - \alpha_1(y_3) + \alpha_1(z_1) + \alpha_1(z_2) - \alpha_1(z_4) + \alpha_2 z_5 + u_1^* \tag{787}$$

$$y_2 = \beta_0 + \beta_1(y_1) - \beta_1(y_3) + \beta_1(z_1) + \beta_2 z_3 + \beta_3 z_6 + \beta_4 z_2 + u_2^* \tag{788}$$

$$y_3 = \gamma_0 + \gamma_1(y_1) + \gamma_1(y_2) + \gamma_1(z_1) + \gamma_1(z_2) + \gamma_2 z_7 + u_3^* \tag{789}$$

- Step 1: Formulate the Likelihood functions for joint errors

There are n observations for each random error terms
$u_{11} u_{12} u_{13} ......u_{1n}; u_{21} u_{22} u_{23} ......u_{2n}; u_{31} u_{32} u_{33} ......u_{3n};$

Joint distribution of errors in this system assumming that there is no contemporaneous correlation across errors

$$P\left(u_{1i}.u_{2i}.u_{3i}\right) = P\left(u_{1i}\right).P\left(u_{2i}\right).P\left(.u_{3i}\right) \tag{790}$$

Assumption on errors
$u_1 \sim N\left(0, \sigma_{u_1}^2\right)$ and $E\left(u_{1i}.u_{1.j}\right) = 0$
$u_2 \sim N\left(0, \sigma_{u_2}^2\right)$ and $E\left(u_{2i}.u_{2.j}\right) = 0$
$u_3 \sim N\left(0, \sigma_{u3}^2\right)$ and $E\left(u_{3i}.u_{3.j}\right) = 0$
Normal density for each error term them is defined as

$$P(u_1) = \left\{\frac{1}{\sigma_{u1}\sqrt{2\pi}}\right\}^n.\left[\exp\left\{-\frac{1}{2}\left(\frac{\sum u_{1i}^2}{\sigma_{u_1}^2}\right)\right\}\right] \tag{791}$$

$$P(u_2) = \left\{\frac{1}{\sigma_{u2}\sqrt{2\pi}}\right\}^n.\left[\exp\left\{-\frac{1}{2}\left(\frac{\sum u_{2i}^2}{\sigma_{u_2}^2}\right)\right\}\right] \tag{792}$$

$$P(u_3) = \left\{\frac{1}{\sigma_{u3}\sqrt{2\pi}}\right\}^n.\left[\exp\left\{-\frac{1}{2}\left(\frac{\sum u_{3i}^2}{\sigma_{u_3}^2}\right)\right\}\right] \tag{793}$$

For the system as a whole the joint distribution of errors is made by combining these three
$P\left(u_{1i}.u_{2i}.u_{3i}\right) = P\left(u_{1i}\right).P\left(u_{2i}\right).P\left(.u_{3i}\right) =$
$\left\{\frac{1}{2\pi\sqrt{2\pi}}\right\}^n.\left\{\frac{1}{\sigma_{u1}\sigma_{u2}\sigma_{u3}}\right\}^{2n}.\left[\exp\left\{-\left(\frac{\sum u_{1i}^2}{2\sigma_{u_1}^2} + \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} + \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2}\right)\right\}\right]$

- Step 2: Derive the likelihood functions of y from the likelihood functions of u

$$u_1 = y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5 \tag{794}$$

$$u_2 = y_2 - \beta_0 - \beta_1(y_1) + \beta_1(y_3) - \beta_1(z_1) - \beta_2 z_3 - \beta_3 z_6 - \beta_4 z_2 \tag{795}$$

$$u_3 = y_3 - \gamma_0 - \gamma_1(y_1) - \gamma_1(y_2) - \gamma_1(z_1) - \gamma_1(z_2) - \gamma_2 z_7 \tag{796}$$

Jacobian determinants of above errors with respect to endogenous variabls

$$|J| = \begin{vmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \frac{\partial u_1}{\partial y_3} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \frac{\partial u_2}{\partial y_3} \\ \frac{\partial u_3}{\partial y_1} & \frac{\partial u_3}{\partial y_2} & \frac{\partial u_3}{\partial y_3} \end{vmatrix} = \begin{vmatrix} 1 & -\alpha_1 & \alpha_1 \\ -\beta_1 & 1 & \beta_1 \\ -\gamma_1 & -\gamma_1 & 1 \end{vmatrix} \tag{797}$$

The likelihood functions for y variables are obtained using the transformation functions as:

$$P\left(y_{1i}.y_{2i}.y_{3i}\right) = P\left(u_{1i}.u_{2i}.u_{3i}\right).\left|\frac{\partial\left(u_{1i}.u_{2i}.u_{3i}\right)}{\partial\left(y_{1i}.y_{2i}.y_{3i}\right)}\right| \tag{798}$$

Using above derivations the likelihood functions

$$
\begin{aligned}
L &= |J|^n \left\{ \frac{1}{2\pi\sqrt{2\pi}} \right\}^n . \left\{ \frac{1}{\sigma_{u1}\sigma_{u2}\sigma_{u3}} \right\}^n . \\
&\left[\exp\left\{ -\left(\frac{\sum u_{1i}^2}{2\sigma_{u_1}^2} + \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} + \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} \right) \right\} \right] \qquad (799)
\end{aligned}
$$

The log likelihood function is given as

$$
lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{\sum u_{1i}^2}{2\sigma_{u_1}^2} - \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} - \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2}
$$

u_i can be expressed in terms of endogenous y and predetermined z variables.

- Step III: First order conditions for maximisation

$\frac{\partial lnL}{\partial \alpha_1} = 0$; $\frac{\partial lnL}{\partial \alpha_2} = 0$; $\frac{\partial lnL}{\partial \beta_1} = 0$; $\frac{\partial lnL}{\partial \beta_2} = 0$; $\frac{\partial lnL}{\partial \beta_3} = 0$; $\frac{\partial lnL}{\partial \beta_4} = 0$; $\frac{\partial lnL}{\partial \gamma_1} = 0$; $\frac{\partial lnL}{\partial \gamma_2} = 0$; $\frac{\partial lnL}{\partial \sigma_{u1}} = 0$; $\frac{\partial lnL}{\partial \sigma_{u2}} = 0$; $\frac{\partial lnL}{\partial \sigma_{u3}} = 0$;

$$
\frac{\partial lnL}{\partial \alpha_i} = n\frac{1}{|J|}.\frac{\partial |J|}{\partial \alpha_i} + \sum \frac{\partial lnL}{\partial u_1}\frac{\partial u_1}{\partial \alpha_i} = 0 \quad for\ i = 1,2 \qquad (800)
$$

$$
\frac{\partial lnL}{\partial \beta_i} = n\frac{1}{|J|}.\frac{\partial |J|}{\partial \beta_i} + \sum \frac{\partial lnL}{\partial u_2}\frac{\partial u_2}{\partial \beta_i} = 0 \ for\ i = 1,2,3,4 \qquad (801)
$$

$$
\frac{\partial lnL}{\partial \sigma_i} = -n\frac{1}{\partial \sigma_{ui}} + \frac{1}{\sigma_{u_i}^3}\sum (u_i)^2 = 0 for\ i = 1,2,3 \qquad (802)
$$

Solving for parameters from these first order conditions is quite involving task becuase of non-linearities of functions in parameters.

For instance consider one case of parameter $\frac{\partial lnL}{\partial \alpha_i} = 0$ for α.

Given the log likelihood function: $lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{\sum u_{1i}^2}{2\sigma_{u_1}^2} - \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} - \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2}$

$\frac{\partial lnL}{\partial \sigma_{u1}} = -\frac{n}{\sigma_{u1}} - \sum u_{1i}^2\left(-\frac{1}{\sigma_{u_1}^3}\right) = -\frac{n}{\sigma_{u1}} + \frac{\sum u_{1i}^2}{\sigma_{u_1}^3} = 0$; This implies $\frac{n}{\sigma_{u1}} = \frac{\sum u_{1i}^2}{\sigma_{u_1}^3}$ or $\sigma_{u_1}^2 = \frac{\sum u_{1i}^2}{n}$.

Similar derviation $\sigma_{u_2}^2 = \frac{\sum u_{2i}^2}{n}$ for$\sigma_{u_2}^2 = \frac{\sum u_{2i}^2}{n}$ and $\sigma_{u3}^2 = \frac{\sum u_{3i}^2}{n}$ for σ_{u3}.

Substitute the values of $\sigma_{u_1}^2, \sigma_{u_2}^2$ and σ_{u3}^2 in the loglikelihood function:

Similar derviation $\sigma_{u_2}^2 = \frac{\sum u_{2i}^2}{n}$ and $\sigma_{u3}^2 = \frac{\sum u_{3i}^2}{n}$.

Substitute the values of $\sigma_{u_1}^2, \sigma_{u_2}^2$ and σ_{u3}^2 in the loglikelihood function:

$$
\begin{aligned}
lnL &= n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{\sum u_{1i}^2}{2\sigma_{u_1}^2} - \frac{\sum u_{2i}^2}{2\sigma_{u_2}^2} - \frac{\sum u_{3i}^2}{2\sigma_{u_2}^2} \\
&= n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{n\sum u_{1i}^2}{2\sum u_{1i}^2} - \frac{n\sum u_{2i}^2}{2\sum u_{2i}^2} - \frac{\sum u_{2i}^2}{2\sum u_{3i}^2}
\end{aligned}
$$

$$
lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{3n}{2}
$$

$lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - n\ln\left(\sigma_{u1}\sigma_{u2}\sigma_{u3}\right) - \frac{3n}{2}$
or writing

$$lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - \frac{n}{2}\ln\left(\sigma^2_{u_1}\sigma^2_{u_2}\sigma^2_{u_3}\right) - \frac{3n}{2}$$

$$lnL = n\ln|J| - n\ln\left(2\pi\sqrt{2\pi}\right) - \frac{n}{2}\left(\ln\sigma^2_{u_1} + \ln\sigma^2_{u_2} + \ln\sigma^2_{u_3}\right) - \frac{3n}{2}$$

Now differentiale lnL wrt to α_1

$$\frac{\partial lnL}{\partial\alpha_1} = n\frac{1}{|J|}.\frac{\partial|J|}{\partial\alpha_1} - \frac{n}{2}\frac{1}{\sigma^2_{u_1}}\frac{\partial\sigma^2_{u_1}}{\partial\alpha_1}$$

substitue $\sigma^2_{u_1} = \frac{\sum u^2_{1i}}{n}$ this to $\frac{\partial lnL}{\partial\alpha_1} = n\frac{1}{|J|}.\frac{\partial|J|}{\partial\alpha_1} - \frac{n}{2}\frac{1}{\sum u^2_{1i}}\frac{\partial\sum u^2_{1i}}{\partial\alpha_1} = n\frac{1}{|J|}.\frac{\partial|J|}{\partial\alpha_1} - \frac{n}{2}\frac{1}{\sum u^2_{1i}}\left(2\sum u_{1i}\right)\frac{\partial u_{1i}}{\partial\alpha_1} = 0$

Evaluate the Jacobian determinant
$|J| = \begin{vmatrix} \frac{\partial u_1}{\partial y_1} & \frac{\partial u_1}{\partial y_2} & \frac{\partial u_1}{\partial y_3} \\ \frac{\partial u_2}{\partial y_1} & \frac{\partial u_2}{\partial y_2} & \frac{\partial u_2}{\partial y_3} \\ \frac{\partial u_3}{\partial y_1} & \frac{\partial u_3}{\partial y_2} & \frac{\partial u_3}{\partial y_3} \end{vmatrix} = \begin{vmatrix} 1 & -\alpha_1 & \alpha_1 \\ -\beta_1 & 1 & \beta_1 \\ -\gamma_1 & -\gamma_1 & 1 \end{vmatrix} = \begin{vmatrix} 1 & \beta_1 \\ -\gamma_1 & 1 \end{vmatrix} + \alpha_1\begin{vmatrix} -\beta_1 & \beta_1 \\ -\gamma_1 & 1 \end{vmatrix} + \alpha_1\begin{vmatrix} -\beta_1 & 1 \\ -\gamma_1 & -\gamma_1 \end{vmatrix}$
$= \left(1 + \gamma_1\beta_1\right) + \alpha_1\left(-\beta_1 + \gamma_1\beta_1\right) + \alpha_1\left(\gamma_1\beta_1 + \gamma_1\right) = 1 + \gamma_1\beta_1 + \alpha_1\left(-\beta_1 + \gamma_1 + 2\gamma_1\beta_1\right)$

Now the partial derivative of Jacobian wrt to α_1 is $\frac{\partial|J|}{\partial\alpha_1} = -\beta_1 + \gamma_1 + 2\gamma_1\beta_1$. Then take u_1 as defined above and differential wrt α_1
$u_1 = y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5$
$\frac{\partial u_{1i}}{\partial\alpha_1} = -y_2 + y_3 - z_1 - z_2 + z_4$

Substitution all these partial derivatives in the $\frac{\partial lnL}{\partial\alpha_1} = n\frac{1}{|J|}.\frac{\partial|J|}{\partial\alpha_1} - \frac{n}{2}\frac{1}{\sum u^2_{1i}}\left(2\sum u_{1i}\right)\frac{\partial u_{1i}}{\partial\alpha_1} = 0$
$\frac{n\left(-\beta_1+\gamma_1+2\gamma_1\beta_1\right)}{1+\gamma_1\beta_1+\alpha_1\left(-\beta_1+\gamma_1+2\gamma_1\beta_1\right)} - \frac{n\sum\left(y_1-\alpha_0-\alpha_1(y_2)+\alpha_1(y_3)-\alpha_1(z_1)-\alpha_1(z_2)+\alpha_1(z_4)-\alpha_2 z_5\right).\left(-y_2+y_3-z_1-z_2-z_4\right)}{\sum\left(y_1-\alpha_0-\alpha_1(y_2)+\alpha_1(y_3)-\alpha_1(z_1)-\alpha_1(z_2)+\alpha_1(z_4)-\alpha_2 z_5\right)^2} = 0$
$\frac{\left(-\beta_1+\gamma_1+2\gamma_1\beta_1\right).\sum\left(y_1-\alpha_0-\alpha_1(y_2)+\alpha_1(y_3)-\alpha_1(z_1)-\alpha_1(z_2)+\alpha_1(z_4)-\alpha_2 z_5\right)^2}{1+\gamma_1\beta_1+\alpha_1\left(-\beta_1+\gamma_1+2\gamma_1\beta_1\right)}$
$= \sum\left(y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5\right).\left(-y_2 + y_3 - z_1 - z_2 - z_4\right)$

- This process should continue for all structural parameters $\alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2$ and for once all these estimates are obtained; then need to evaluate the variances of errors $\sigma^2_{u_1}, \sigma^2_{u_2}$ and $\sigma^2_{u_3}$.
- After having estimates of all parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2$ and $\sigma^2_{u_1}, \sigma^2_{u_2}$ and $\sigma^2_{u_3}$ you can evaluate the Likelihood function.
- Structural parameters should be obtained by solving these first order conditions as illustrated in simple example above.

References

[1] Amemiya T (1975)The nonlinear limited-information maximumlikelihood estimator and the modified nonlinear two-stage least-squares estimator, Journal of Econometrics 3, 375-386.

[2] Amemiya T (1977) The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous EquationEconometrica, 45, 4, 955-968

[3] Anderson T. W. (1974) An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System, Journal of the American Statistical Association, 69, 346, 565-573

[4] Aït-Sahalia Y (2002),Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach, Econometrica, 70, 1 223–262

[5] Blundell R. W.K. Newey and T. Persson (2006) Advances in Economics and Econometrics, Econometric Society Monograph, Vol. II and III, Cambridge University Press.

[6] Chesher A and R. Spady (1991) Asymptotic Expansions of the Information Matrix Test Statistics, Econometrica, 59, 3, 787-815

[7] den Haan W.J. and A Marcet (1990) Solving the stochastic growth model by parameterising expectations, *Jounral of Business and Economic Statistics*, 8:1:31-34.

[8] Hendry D. F. and J. A. Doornik (1994) Modelling linear dynamic econometric systems, Scottish Journal of Political Economy, 41,1,1-33

[9] Durham, G B,Gallant, A R.;Ait-Sahalia, Y and Brandt, M. W (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, Journal of Business & Economic Statistics; Jul 2002; 20, 3

[10] Enders W. (2010) Applied Econometric Time Series, 3rd ed., Willey.

[11] Greene W. (2008) Econometric Analysis, Fifth Edition, Prentice Hall.

[12] Fukac and A Pagan (2010) Limited Information Estimation and Evaluation of DSGE Models, J. Appl. Econ. 25: 55–70

[13] Hamilton James D. (1994) Time Series Analysis, Princeton.

[14] Harvey A. C. and G. D. A. Phillips (1979) Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances Biometrika, 66, 1, 49-58

[15] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[16] Hartley M. J. and P. Mallela (1977) The Asymptotic Properties of a Maximum Likelihood Estimator for a Model of Markets in Disequilibrium, Econometrica, 45, 5, 1205-1220

[17] Hendry (1995) Dynamic Econometric Theory, Oxford.

[18] Hendry D. F. (1971) Maximum Likelihood Estimation of Systems of Simultaneous Regression Equations with Errors Generated by a Vector Autoregressive Process, International Economic Review, 12, 2, 257-272

[19] Hillier G and M Armstrong (1999) The Density of the Maximum Likelihood Estimator, Econometrica, 67, 6, 1459-1470

[20] Karlson and Skoglund (2000) Maximum-likelihood based inference in the two-way random effects, ESEM.

[21] Kawakatsu Hiroyuki (2007) Numerical integration-based Gaussian mixture filters for maximum likelihood estimation of asymmetric stochastic volatility models, Econometrics Journal 10, 342–358.

[22] Klein L R (1956) A Textbook of Econometrics, Row Petterson and Company, New York

[23] Koutsoyiannis A. (1973) Theory of Econometrics, Second Edition, McMillan.

[24] Lee Lung-fei (1993) Asymptotic Distribution of the Maximum Likelihood Estimator for a Stochastic Frontier Function Model with a Singular Information Matrix, Econometric Theory, 9, 3, 413-430

[25] Lüutkepohl Helmut, Pentti Saikkonen, Carsten Trenkler ( ) Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process, Econometrics Journal, 4, 2, 287–310.

[26] Maddala G.S. (2003) Introduction to Econometrics, Cambridge.

[27] Nielsen M Ø (2004) Efficient inference in multivariate fractionally integrated time series models,Econometrics Journal, 7, 63–97

[28] Phillips PCB and B. M. Pötscher (1993) Efficiency of Maximum Likelihoo, Econometric Theory, 9, 3,534-536

[29] PhillipsP. C. B.(1976) The Iterated Minimum Distance Estimator and the Quasi-Maximum Likelihood Estimator, Econometrica, 44, 3, 449-460

[30] Pearson K (1936) Method of Moments and Method of Maximum Likelihood Source: Biometrika, 28, 1/2,34-59

[31] Ruud P. A. (2000) An Introduction to Classical Econometric Theory, Oxford.

[32] Sweeting T. J. (1980) Uniform Asymptotic Normality of the Maximum Likelihood Estimator,The Annals of Statistics, 8, 6, 1375-1381

[33] Wald A (1943) Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large,Transactions of the American Mathematical Society, 54, 3, 426-482

3.2 Spectral Analysis

Time series Y_t can be expressed in time domain and frequency domain (See Hamilton (1994).

Time domain:

$$Y_t = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_t \tag{803}$$

Frequency domain or spectral analysis

$$Y_t = \mu + \int_0^{\pi} \alpha(\omega) \cos(\omega t)\, d\omega + \int_0^{\pi} \delta(\omega) \sin(\omega t)\, d\omega \tag{804}$$

Spectral Analysis

A covariance stationary process $\{Y_t\}_0^\infty$ with $E(Y_t) = \mu$ and j th covariance $E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$.

Autocovariance generating function

$$g_y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j \tag{805}$$

$$z^j = e^{-i\omega} \quad i = \sqrt{-1} \tag{806}$$

Here ω is a frequency. It is assumed that these autocovariances γ_j are absolutely summable.

Population spectrum of Y_t is given by

$$S_y(\omega) = \frac{1}{2\pi} g_y(z) = \frac{1}{2\pi} g_y\left(e^{-i\omega}\right) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j \left(e^{-i\omega j}\right) \tag{807}$$

Spectral Analysis

From De Moivre's theorem $e^{-i\omega j} = \cos(\omega j) - i \operatorname{Si} n(\omega j)$

$$S_y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j \left(\cos(\omega j) - i \operatorname{Si} n(\omega j)\right) \tag{808}$$

Note the trigonometric rules $\cos(0) = 1; \sin(0) = 0; \cos(-\theta) = \cos\theta; \sin(-\theta) = -\sin(\theta)$; $\sin = \cos\theta$ $\frac{\partial \sin\theta}{\partial\theta} = \cos\theta$; $\frac{\partial \cos\theta}{\partial\theta} = -\sin\theta$;

$$S_y(\omega) = \tfrac{1}{2\pi}\gamma_0 \left(\cos(0) - i \operatorname{Si} n(0)\right) + \tfrac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j \left(\cos(\omega j) + \cos(-\omega j) - i \operatorname{Si} n(\omega j) - i \operatorname{Si} n(-\omega j)\right)$$

For the covariance stationary proces $\gamma_j = \gamma_{-j}$

Spectral Analysis

$$S_y(\omega) = \frac{1}{2\pi}\left[\gamma_0 \cos(0) + 2\sum_{j=-\infty}^{\infty} \gamma_j \cos(\omega j)\right] \tag{809}$$

$S_y(\omega)$ is a continous real valued function of ω.

Prove that population spectral representation of $ARMA(p,q)$ process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_p \varepsilon_{t-q} \tag{810}$$

is given by

$$\begin{aligned} S_y(\omega) \quad &= \quad \frac{\sigma^2}{2\pi} \frac{\left(1 + \theta_1 e^{-i\omega} + \theta_2 e^{-i2\omega} + ... + \theta_q e^{-iq\omega}\right)}{\left(\left(1 - \phi_1 e^{-i\omega} + \phi_2 e^{-i2\omega} + ... + \phi_{2p} e^{-ip\omega}\right)\right)} \\ &\times \frac{\left(1 + \theta_1 e^{i\omega} + \theta_2 e^{i2\omega} + ... + \theta_q e^{iq\omega}\right)}{\left(\left(1 - \phi_1 e^{i\omega} + \phi_2 e^{i2\omega} + ... + \phi_{2p} e^{ip\omega}\right)\right)} \end{aligned} \tag{811}$$

3.2.1 Brownian Motion

Consider a random walk

$$y_t = y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \text{ iid } N(0,1) \tag{812}$$

$$y_t = \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 + \epsilon_5 + ... + \epsilon_t \tag{813}$$

change in value of y_t between periods t and s

$$\begin{aligned} y_t - y_s &= \epsilon_{t+1} + \epsilon_{t+2} + \epsilon_{t+3} + \epsilon_{t+4} + \epsilon_{t+5} + ... + \epsilon_s \\ &\sim N(0,(s-t)) \end{aligned} \tag{814}$$

Divide the error
$y_t - y_{t-1} = \varepsilon_t$

$$\varepsilon_t = e_{1t} + e_{2t} \tag{815}$$

Now divite the time interval in infinitely small sections

$$\varepsilon_t = e_{1t} + e_{2t} + ... + e_{Nt} \sim N\left(0, \frac{1}{N}\right) \tag{816}$$

when N $\longrightarrow \infty$ it is called a Brownian motion, $W(t)$. This is continous stochastic function and has folliwng properties

- $W(0). = 0$
- For any dates $0 \leq t_1 \leq t_2 \leq ... \leq t_k \leq 1$ the changes
- $[W(t_2) - W(t_1)], [W(t_3) - W(t_2)], ..., [W(t_k) - W(t_{k-1})]$ are independent multivariate Gaussian with $[W(s) - W(t)]\ \tilde{}N(0,(s-t))$
- Any realisation of $W(t)$ is continous in$t$with probability 1.

Other continous time process can be generated from standard Browning motions, as:

$$Z(t) = \sigma W(t) \sim N\left(0, \sigma^2 t\right) \tag{817}$$

- Hamilton James D. (1994) Time Series Analysis, Princeton.
- For issues see http://www.weforum.org/

4 L3: Time Series Analysis

A time series is a data generating process and it is characterised by its mean variance and autocovariance. Trends, cycles, seasonality and irregular components of a time series are often represented in terms of autoregressive or moving average models or combinations of these. What is the right order of autoregressive or moving average component and combination of these two and what sorts of autocorrelation coefficients are required for stationary series and how can non-stationary series be made stationary and used for economic analysis is one of the major topic in univariate time series analyses.

Multivariate vector autoregressive (VAR) and simultaneous equation models are used to analysis of economic systems. These involve a number of equations with various time series variables. What is the long run relationship among variables? How do adjustments occur towards the long run whenever these process deviate from the long run equilibrium? How can economic variables cbe coitegrated are discussed in multivariate time series analyses? Works of Klein (1956), Box and Jenkins (1976), Hamilton(1994), Harvey (1976), Dickey and Fuller (1979), Hendry (1995), Engle (1982), Engle and Granger (1987), Phillips (1987), Stock and Watson (2002), Nelson and Plosser (1982), Pagan and Wickens (1989), Pyndick and Rubinfeld (1998), Wooldridge (1994) , Enders (2010) and others have build the body of knowledge in econometrics.

A series of macro time series models have appeared in the literature. Such modelling exercises have become more intensive in recent years as more information on the major macroeconomic variables have become available. These have become even more relevant in torbullent times in recent years along with availability, regularity and quality of high frequency data sets. These models broadly can be divided into four main categories. First, the Keynesian IS-LM model for closed or open economies are based on structural equations to explain demand sides of the economy assuming a fixed supply in the short run. When supply shocks, such as the higher oil prices hit economies around the world in 1970s scepticism increased on the outcome of the demand determined solutions of the Keynesian model as they were inconsistent with the stagflationary experience of the many advanced economies. Mainly, three other alternative models have been proposed to explain the emerging realities of these economies. One approach is to use time series of a particular variable for forecasting for the short run. It is done either by single equation model such as the ARIMA, or multiple equation model such as the vector autoregression (VAR) models or structural cointegration (CVAR) models. Another approach is to use small scale macro models with rational expectation or/and micro foundation. Finally there are infinite horizon dynamic general equilibrium models for the decentralised markets with clear focus on the real side of the economy. New classical macro economists use market clearing models with stochastic shocks to technology or fiscal policy in explaining the evolution of the economy over time. New Keynesians, on the other hand, reinstate the Keynesian conclusions using more sophisticated modelling technology that takes account of explicit optimisation by households and firms as developed by new classical economists. There are excellent surveys on macroeconometric modelling in the literature (Wallis (1989), Pagan and Wickens (1989)). As they account how many authors have contributed to the macroeconometric modelling and forecasting in the UK since 1969. Klein (1968), Sims (1967), Cairncorss (1969), Hendry (1974), Granger and Newbold (1975), Ash and Smyth (1978), Doan, Litterman and Sim (1984), Burns (1986) have either used simultaneous model based approach or the time series approach for forecasting. (see Holly and Weal (2000) for some recent development in this area). Though forecasting has been in practice quite widely both by the government and the private sector appreciable difference remains across various forecasting groups such as the London Business School (LBS), National

Institute of Social and Economic Research (NISER, NIGEM and NIDEM), Liverpool University Research Group (LPL) and Cambridge University group (CUBS). Such experience is also evident in the US models such as DRI, WHARTON or TAYLOR or the OECD models. Despite that model based macroeconomic forecasts are found to have large prediction errors. Clements and Hendry (2002) warrant for improved procedures so that the model based forecasts to be more accurate than simple and plain extrapolative forecasts. More recently Garratt (et.al.) illustrate a structural cointegrating VAR approach to macroecometric modelling while comparing developments on each of the above four models. Al most all of the above models lack dynamic general equilibrium framework for decentralised markets where resources are allocated by prices that are determined by underlying forces of demand and supply in the economy (see Bhattarai (2011) for illustration of one of these models).

ARMA: Box and Jenkins (1976), Lütkepohl (1984), Evans and Honkapohja (1986), Campos (1986), McCabe and Leybourne (1998), Pesaran, Shin, Smith (2001), Ling and McAleer (2003) Bailey (2007) Pan, Wang and Yao (2007)

Time Series: Klein (1956), Box and Jenkins (1976), Hamilton(1994), Harvey (1976), Dickey and Fuller (1979), Hendry(1995), Engle (1982), Engle and Granger (1987), Phillips (1987), Stock and Watson (2002), Nelson and Plosser (1982), Pagan and Wickens (1989), Pyndick and Rubinfeld (1998), Wooldridge (1994), Enders (2010), Sims (1980), Beveridge and Nelson (1981),Pesaran (1982) Johansen (1988), Baltagi (1992), Pesaran and Smith (1995), Garratt, Lee, Phillips (2003) Pesaran and Shin (2003), Hendry (1997),Mills, Pelloni, Zervoyianni (1995), Nelson (1987), Stock and Watson (2001)

ARCH-GARCH: Engle (1982), Engle and Granger (1987), Bollerslev (1986, 1990) Engle (2001, 2002), Engle and Kroner (1995), Baillie and Myers(1991) Kleibergen and Van Dijk(1993) , Lumsdaine (1995), Bauwens, Laurent and Rombouts (2006),Huang, Wang and Yao (2008)

4.0.2 Definition of a time series

Time series is the data generating process $\{y_t\}_{-\infty}^{\infty} = \left\{ \begin{array}{c} y_{-\infty}.....y_{-1}.y_0.y_1.y_2.... \\ ....y_T.y_{T+1}.y_{T+1}..... \end{array} \right\}$

A time series consists of trend, cycle, season and irregular component

$$Y = T \times C \times S \times I \tag{818}$$

In a simple method the moving average gives $T \times C$ components and is used to isolate the $S \times I$ components. For instance for a 12 monthly moving average

$$\overline{Y}_i = \frac{1}{12}\left(Y_1 + Y_2 + ... + Y_{12}\right) \tag{819}$$

$$S \times I = \frac{T \times C \times S \times I}{T \times C} = \frac{Y_i}{\overline{Y}_i} = z_t \tag{820}$$

Now to isolate the Irregular component I from $S \times I$ take out the seasonal elements from z_t assuming monthly data for 5 years (60 observations) compute the seasonal indices as following:

$$Month1: \ \overline{z}_1 = \frac{1}{5}\left(z_1 + z_{13} + z_{25} + z_{39} + z_{48}\right) \tag{821}$$

$$Month2: \ \overline{z}_2 = \frac{1}{5}\left(z_2 + z_{14} + z_{26} + z_{40} + z_{49}\right) \tag{822}$$

$$Month3: \ \overline{z}_3 = \frac{1}{5}\left(z_3 + z_{15} + z_{26} + z_{41} + z_{50}\right) \tag{823}$$

....

$$Month11: \ \overline{z}_{11} = \frac{1}{5}\left(z_{11} + z_{23} + z_{35} + z_{47} + z_{59}\right) \tag{824}$$

$$Month12: \ \overline{z}_{12} = \frac{1}{5}\left(z_{12} + z_{24} + z_{36} + z_{46} + z_{60}\right) \tag{825}$$

Deseasonalisation of data $Y_i^d = \frac{Y_i}{\overline{z}_i}$ and irregular component should be $i = \frac{z_t}{\overline{z}_i}$.

Trends: Simple extrapolation

$$Y_t = c_1 + c_2 t \tag{826}$$

Exponential growth

$$Y_t = Ae^{rt} \tag{827}$$

Autoregressive model

$$Y_t = c_1 + c_2 Y_{t-1} \tag{828}$$

Log trend

$$\ln\left(Y_t\right) = c_1 + c_2 \ln\left(Y_{t-1}\right) \tag{829}$$

Quadratic trends:

$$Y_t = c_1 + c_2 t + c_3 t^2 \tag{830}$$

Logistic trend:

$$Y_t = \frac{1}{k + bt} \quad b > 1 \tag{831}$$

$$Y_t = e^{k_1 - \frac{k_2}{t}} \tag{832}$$

$$ln\left(Y_t\right) = k_1 - \frac{k_2}{t} \tag{833}$$

auto lagged with declining weights $\alpha < 1$

$$Y_t = \alpha Y_{t-1} + \alpha\left(1-\alpha\right)Y_{t-2} + \alpha\left(1-\alpha\right)^2 Y_{t-2} + ... + \alpha\left(1-\alpha\right)^n Y_{t-2} \tag{834}$$

Forecasting forward with these models is obvious.

4.0.3 Time Series models

Time series models aim to explain the data generating process for $\{y_t\}_{-\infty}^{\infty} = \{y_{-\infty}.....y_{-1}.y_0.y_1.y_2........y_T.y_{T+1}.y_{T+1}.....\}$

Simplest of these is a trend model

$$Y_t = \beta t + \varepsilon_t \tag{835}$$

with mean $E(Y_t) = \beta t$ and variance $E\left(Y_t - \beta t\ \right)^2 = E\left(\varepsilon_t\right)^2 = \sigma_\varepsilon^2$

Or it could have been just a constant plus a Gaussian white noise $\varepsilon_t \sim N\left(0, \sigma^2\right)$ as:

$$Y_t = \mu + \varepsilon_t \tag{836}$$

with mean $E(Y_t) = \mu$ and variance $E\left(Y_t - \mu\ \right)^2 = E\left(\varepsilon_t\right)^2 = \sigma_\varepsilon^2$

Autocovariance of $\{y_t\}_{-\infty}^{\infty}$ for I realisations is

$$\gamma_{tj} = E\left(Y_t - \mu\ \right) E\left(Y_{t-j} - \mu\ \right) = E\left(\varepsilon_t\right) E\left(\varepsilon_{t-j}\right) = 0 \quad for\ j \neq 0 \tag{837}$$

Stationarity: when neither mean μ nor the autocovariance γ_{ij} depend on time t then the Y_t is covariance stationary or weakly stationary.

$$E(Y_t) = \mu\ for\ \forall\ t \tag{838}$$

$$E\left(Y_t - \mu\right) E\left(Y_{t-j} - \mu\right) = \gamma_j \quad for \text{ any } t \ \ and\ j = \left\{ {}^{\sigma_\varepsilon^2\ for\ j=0}_{0\ for\ j\neq 0} \right. \tag{839}$$

For instance (836) is stationary while (835) not covariance stationary because its mean βt is function of time.

If the process is stationary γ_j is the same for any value of t $\gamma_j = \gamma_{-j}$

$$\begin{aligned} \gamma_j &= E\left(Y_{t+j} - \mu\right) E\left(Y_{(t+j)-j} - \mu\right) = E\left(Y_{t+j} - \mu\right) E\left(Y_t - \mu\right) \\ &= E\left(Y_t - \mu\right) E\left(Y_{t+j} - \mu\right) \ = \gamma_{-j} \end{aligned} \tag{840}$$

Ergodicity A covariance process is ergodic for the mean if the mean for certain observations $\overline{y} \equiv \frac{1}{T}\sum_{t=1}^{T} y_t^{(1)}$ converge in probability to $E(Y_t)$ as $T \longrightarrow \infty$. It is ergodic in mean if

$$\sum_{t=1}^{T} \left|\gamma_j\right| < \infty \tag{841}$$

A covariance process is ergodic for the second moment if

$$\left[\frac{1}{T-j}\right] \sum_{t=1}^{T} \left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right) \xrightarrow{p} \gamma_j \tag{842}$$

Stationarity and ergodicity are same in most applications, except few exceptions.

4.1 Moving Average Process

First order moving average process: MA(1)

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \tag{843}$$

Here μ, θ could be any constant. Y_t is constructed from two most recent realisation of white noise $\varepsilon_t \sim N\left(0, \sigma^2\right)$.

Mean of Y_t

$$E\left(Y_t\right) = E\left(\mu + \varepsilon_t + \theta\varepsilon_{t-1}\right) = \mu + E\left(\varepsilon_t\right) + \theta E\left(\varepsilon_{t-1}\right) = \mu \tag{844}$$

Variance of Y_t

$$\begin{aligned} E\left(Y_t - \mu\right)^2 &= \gamma_0 = E\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)^2 = E\left(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_t^2\right)^2 \\ &= \sigma^2 + 0 + \theta^2\sigma^2 = \sigma^2\left(1 + \theta^2\right) \end{aligned} \tag{845}$$

MA(1) has a memory of only 1 period.

First autocovariance:

$$\begin{aligned} E\left(Y_t - \mu\right)E\left(Y_{t-1} - \mu\right) &= \gamma_1 \qquad (846) \\ &= E\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)\left(\varepsilon_{t-1} + \theta\varepsilon_{t-2}\right) \\ &= E\left(\varepsilon_t\varepsilon_{t-1} + \theta\varepsilon_t\varepsilon_{t-2} + \theta\varepsilon_{t-1}\varepsilon_{t-1} + \theta\varepsilon_{t-1}\theta\varepsilon_{t-2}\right) \\ &= 0 + 0 + \theta\sigma^2 + 0 = \theta\sigma^2 \end{aligned}$$

Higher autocovariance:

$$\begin{aligned} E\left(Y_t - \mu\right)E\left(Y_{t-j} - \mu\right) &= \gamma_j = E\left(\varepsilon_t + \theta\varepsilon_{t-1}\right)\left(\varepsilon_{t-j} + \theta\varepsilon_{t-j-1}\right) = 0 \\ for\ j &> 1 \end{aligned} \tag{847}$$

MA(1) process is ergotic for all moments.

Autocorrelation

$$\rho_j = \frac{cov\left(Y_t, Y_{t-j}\right)}{\sqrt{var(Y_t)}\sqrt{Y_{t-j}}} = \frac{\gamma_j}{\gamma_0} \tag{848}$$

$$\rho_1 = \frac{\theta\sigma^2}{\sqrt{\sigma^2\left(1 + \theta^2\right)}\sqrt{\sigma^2\left(1 + \theta^2\right)}} = \frac{\theta}{\left(1 + \theta^2\right)} < 1 \tag{849}$$

MA(q) process

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q} \tag{850}$$

Mean

$$E\left(Y_t\right) = E\left(\mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}\right) = \mu \tag{851}$$

Variance

$$\begin{aligned} E\left(Y_t-\mu\right)^2 &= \gamma_0=E\left(\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+....+\theta_q\varepsilon_{t-q}\right)^2 \\ &= \sigma^2+\theta_1^2\sigma^2+\theta_2^2\sigma^2+..+\theta_q^2\sigma^2 \\ &= \left(1+\theta_1^2+\theta_2^2+..+\theta_q^2\right)\sigma^2 \end{aligned} \tag{852}$$

MA(q) has a memory up to q periods.
Autocovariance for j = 1, 2, q;

$$\begin{aligned} E\left(Y_t-\mu\right)E\left(Y_{t-1}-\mu\right) &= \gamma_1 \\ &= E\left(\mu+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+....+\theta_q\varepsilon_{t-q}\right) \\ &\quad \left(\mu+\varepsilon_{t-j}+\theta_1\varepsilon_{t-j-1}+\theta_2\varepsilon_{t-j-2}+....+\theta_q\varepsilon_{t-j-q}\right) \\ &= E\left(\theta_j\varepsilon_{t-j}^2+\theta_{j+1}\theta_1\varepsilon_{t-j-1}^2+\theta_{j+2}\theta_2\varepsilon_{t-j-2}^2+...+\theta_{j+q}\theta_q\varepsilon_{t-j-q}^2\right) \\ &= \sigma^2\left(\theta_j+\theta_{j+1}\theta_1+\theta_{j+2}\theta_2+...+\theta_{j+q}\theta_q\right) \end{aligned} \tag{853}$$

$$\gamma_j = \{^{\sigma^2(\theta_j+\theta_{j+1}\theta_1+\theta_{j+2}\theta_2+...+\theta_{j+q}\theta_q) \quad for\ j=1,2,...q}_{0\ for\ j>q} \tag{854}$$

For MA(2) process

$$\begin{gathered} \gamma_0=\left(1+\theta_1^2+\theta_2^2\right)\sigma^2 \\ \gamma_1=\left(\theta_1+\theta_2\theta_1\right)\sigma^2 \\ \gamma_2=\left(\theta_2\right)\sigma^2 \\ \gamma_3=\gamma_4=\gamma_5=....=0 \end{gathered}$$

Similarly autocovariance function is vanishes after q lags for the MA(q) process.

$$Y_t=\mu+\sum_{j=1}^{p}\theta_j\varepsilon_{t-j} \tag{855}$$

Infinite order moving average process MA(∞)

$$Y_t=\mu+\sum_{j=1}^{\infty}\psi_j\varepsilon_{t-j} \tag{856}$$

1) For Gausian white noise MA(∞) is ergodic $\sum_{j=1}^{\infty}\psi_j^2<\infty$. Its covariances are absolutely summable $\sum_{j=1}^{\infty}\left|\gamma_j\right|<\infty$. Prove it.

4.2 Autoregressive Process

First order AR(1)

$$Y_t=c+\phi Y_{t-1}+\varepsilon_t \tag{857}$$

with white noise $\varepsilon_t \sim N\left(0, \sigma^2\right)$.
If $\phi \geqslant 1$ time series Y_t accumulates ϵ_t shocks and the Y_t is not covariance stationary.
If $\phi < 1$ impact of ϵ_t shocks die out and Y_t is covariance stationary and Y_t has stable solution as

$$\begin{aligned} Y_t &= (c+\varepsilon_t) + \phi\,(c+\varepsilon_{t-1}) + \phi^2\,(c+\varepsilon_{t-2}) + \phi^3\,(c+\varepsilon_{t-3}) + .. + .. \\ &= \frac{c}{1-\phi} + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + ..+ \end{aligned} \tag{858}$$

This is in fact MA(∞) process. When $|\phi| < 1$

$$\sum_{j=1}^{\infty} \left|\psi_j\right| = \sum_{j=1}^{\infty} |\phi|^j = \frac{1}{1-\phi} \tag{859}$$

Thus MA(∞) representation exists for any AR(1) process.
AR(1) process is ergodic
Mean of AR(1)

$$E\left(Y_t\right) = E\left(\frac{c}{1-\phi} + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + ..+\right) \tag{860}$$

$$E\left(Y_t\right) = \frac{c}{1-\phi} + 0 + 0 + +..+ \tag{861}$$

$$\mu = \frac{c}{1-\phi} \tag{862}$$

Variance of AR(1)

$$\begin{aligned} E\left(Y_t-\mu\right)^2 &= \gamma_0 = E\left(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + ..+\right)^2 \\ &= E\left(1 + \phi^2\varepsilon_{t-1} + \phi^4 + \phi^6 + ..+\right)\sigma^2 \end{aligned} \tag{863}$$

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2} \tag{864}$$

$$c = \mu\,(1-\phi) \tag{865}$$

$$Y_t = \mu\,(1-\phi) + \phi Y_{t-1} + \varepsilon_t \tag{866}$$

$$Y_t - \mu = \phi\left(Y_{t-1} - \mu\right) + \varepsilon_t \tag{867}$$

$$E\left(Y_t-\mu\right)^2 = \phi^2 E\left(Y_{t-1}-\mu\right)^2 + 2\phi E\left(Y_{t-1}-\mu\right)\varepsilon_t + E\left(\varepsilon_t^2\right) \tag{868}$$

$E\left(Y_{t-1}-\mu\right)\varepsilon_t = 0$

$$\gamma_0 = \phi^2\gamma_0 + 0 + \sigma^2 \tag{869}$$

$$\gamma_0 = \frac{\sigma^2}{1-\phi^2} \tag{870}$$

Autocovariane of AR(1)

$$\begin{aligned}
\gamma_j &= E\left(Y_t - \mu\right) E\left(Y_{t-j} - \mu\right) = \\
& E\left(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + .. + \phi^j\varepsilon_{t-j} + \phi^{j+1}\varepsilon_{t-j-1} .. + ..\right) \\
& \left(\varepsilon_{t-j} + \phi\varepsilon_{t-j-1} + \phi^2\varepsilon_{t-j-2} + \phi^3\varepsilon_{t-j-3} + ..+\right) && (871) \\
&= \left(\phi^j + \phi^{j+2} + \phi^{j+4} + ...\right)\sigma^2 && (872)
\end{aligned}$$

$$\begin{aligned}
\gamma_1 &= E\left(Y_t - \mu\right) E\left(Y_{t-1} - \mu\right) = \phi E\left(Y_{t-1} - \mu\right)\left(Y_{t-1} - \mu\right) && (873) \\
+E\left(Y_{t-1} - \mu\right)\varepsilon_t &= \phi\gamma_{1-1} = \phi\gamma_0 && (874)
\end{aligned}$$

$$\begin{aligned}
\gamma_j &= E\left(Y_t - \mu\right) E\left(Y_{t-j} - \mu\right) = \phi E\left(Y_{t-1} - \mu\right)\left(Y_{t-j} - \mu\right) && (875) \\
+E\left(Y_{t-j} - \mu\right)\varepsilon_t &= \phi\gamma_j = \phi^j\gamma_0 && (876)
\end{aligned}$$

Autocorrelation in AR(1)

$$\rho_j = \frac{cov\left(Y_t, Y_{t-j}\right)}{\sqrt{var(Y_t)}\sqrt{Y_{t-j}}} = \frac{\gamma_j}{\gamma_0} = \frac{\phi^j\gamma_0}{\gamma_0} = \phi^j \tag{877}$$

Second order autocorrelation process AR(2)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{878}$$

$$Y_t(1 - \phi_1 L - \phi_2 L^2) = \psi(L)Y_t = c + \varepsilon_t \tag{879}$$

Let $\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1}$

A second order difference equation is stable if the roots of the polynomial are outside the unit circle or eigen values lie inside the unit circle.

Dividing both sides by

$$Y_t = \psi(L)c + \psi(L)\varepsilon_t \tag{880}$$

MA(∞) representation of AR(2) process.

$$\psi(L)c = \frac{c}{(1 - \phi_1 - \phi_2)} \tag{881}$$

And

$$\sum_{j=1}^{\infty} \left|\psi_j\right| < \infty \tag{882}$$

$$\sum_{j=1}^{\infty} \left|\psi_j\right| < \infty \tag{883}$$

Alternatively

$$E\left(Y_t\right) = c + \phi_1 E\left(Y_{t-1}\right) + \phi_2 E\left(Y_{t-2}\right) + \varepsilon_t \tag{884}$$

$$\mu = c + \phi_1 \mu + \phi_2 \mu + 0 \Longrightarrow \mu = \frac{c}{\left(1 - \phi_1 - \phi_2\right)} \tag{885}$$

Variance of AR(2)

$$c = \mu\left(1 - \phi_1 - \phi_2\right) \tag{886}$$

$$Y_t = \mu\left(1 - \phi_1 - \phi_2\right) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{887}$$

$$Y_t - \mu = \phi_1\left(Y_{t-1} - \mu\right) + \phi_2\left(Y_{t-2} - \mu\right) + \varepsilon_t \tag{888}$$

Multiplying both sides by $(Y_t - \mu)$ and taking expectations

$$\begin{aligned} \gamma_0 &= E\left[\left(Y_t - \mu\right)\left(Y_t - \mu\right)\right] = \phi_1 E\left(Y_{t-1} - \mu\right)\left(Y_t - \mu\right) \\ &\quad + \phi_2 E\left(Y_{t-2} - \mu\right)\left(Y_t - \mu\right) + E\left[\varepsilon_t\left(Y_t - \mu\right)\right] \end{aligned} \tag{889}$$

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2 \tag{890}$$

$$E\left[\varepsilon_t\left(Y_t - \mu\right)\right] = \sigma^2$$
$$E\left[\varepsilon_t\left(c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t - \mu\right)\right] = \sigma^2$$

For autocovariance multiply by $(Y_{t-j} - \mu)$ and take the expectation

$$\begin{aligned} \gamma_j &= E\left[\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right)\right] = \phi_1 E\left[\left(Y_{t-1} - \mu\right)\left(Y_{t-j} - \mu\right)\right] \\ &\quad + \phi_2 E\left[\left(Y_{t-2} - \mu\right)\left(Y_{t-j} - \mu\right)\right] + E\left[\varepsilon_t\left(Y_{t-j} - \mu\right)\right] \end{aligned} \tag{891}$$

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad \text{for j = 1,2} \tag{892}$$

The autocovariances follow the second order difference equation as Y_t . AR(2) process is stable if ϕ_1 and ϕ_2 lie inside the unit triangle.

Autocorrelation of AR(2) are found by dividing the autocovariance by variance $\rho_j = \frac{\gamma_j}{\gamma_0}$ Yule-Walker equations

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1 \frac{\gamma_{j-1}}{\gamma_0} + \phi_2 \frac{\gamma_{j-2}}{\gamma_0} \tag{893}$$

$$\rho_1 = \phi_1 + \phi_2\rho_1; \qquad \rho_1 = \frac{\phi_1}{1-\phi_2} \tag{894}$$

$$\rho_2 = \phi_1\frac{\gamma_{2-1}}{\gamma_0} + \phi_2\frac{\gamma_{2-2}}{\gamma_0} = \phi_1\rho_1 + \phi_2 \tag{895}$$

Pth order autocorrelation process: AR(p)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \tag{896}$$

AR(p) is process stable if the roots of the polynomial $1 - \phi_1 z - \phi_2 z^2 - .. - \phi_p z^p = 0$ lie outside the unit circle.

$$Y_t(1 - \phi_1 L - \phi_2 L^2 - .. - \phi_p L^p) = \psi(L)Y_t = c + \varepsilon_t \tag{897}$$

Dividing both sides by MA(∞) representation of AR(p) process.

$$Y_t = \psi(L)c + \psi(L)\varepsilon_t \tag{898}$$

$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - .. - \phi_p L^p)^{-1}$

$$\psi(L)c = \frac{c}{\left(1 - \phi_1 - \phi_2 - .... - \phi_p\right)} \tag{899}$$

Alternatively

$$E\left(Y_t\right) = c + \phi_1 E\left(Y_{t-1}\right) + \phi_2 E\left(Y_{t-2}\right) + ... + \phi_p E\left(Y_{t-p}\right) + \varepsilon_t \tag{900}$$

$$\mu = c + \phi_1\mu + \phi_2\mu + ... + \phi_p + 0 \Longrightarrow \mu = \frac{c}{\left(1 - \phi_1 - \phi_2 - ... - \phi_p\right)} \tag{901}$$

For autocovariance multiply by $(Y_{t-j} - \mu)$ and take the expectation

$$\begin{aligned}\gamma_j \quad = \quad & E\left[\left(Y_t - \mu\right)\left(Y_{t-j} - \mu\right)\right] = \phi_1 E\left[\left(Y_{t-1} - \mu\right)\left(Y_{t-j} - \mu\right)\right] \\ & +\phi_2 E\left[\left(Y_{t-2} - \mu\right)\left(Y_{t-j} - \mu\right)\right] + \\ & ... + \phi_p E\left[\left(Y_{t-p} - \mu\right)\left(Y_{t-j} - \mu\right)\right] + E\left[\varepsilon_t\left(Y_{t-j} - \mu\right)\right]\end{aligned} \tag{902}$$

$$\gamma_j = \{^{\phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + ... + \phi_p\gamma_{j-p} \text{for j = 1,2,...}}_{\phi_1\gamma_1 + \phi_2\gamma_2 + ... + \phi_p\gamma_p \text{for j = 0}} \tag{903}$$

Yule-Walker equations

$$\rho_j = \phi_1\frac{\gamma_{j-1}}{\gamma_0} + \phi_2\frac{\gamma_{j-2}}{\gamma_0} + ... + \phi_p\frac{\gamma_{j-p}}{\gamma_0} \tag{904}$$

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + ... + \phi_p\rho_{j-p} \quad \text{for j = 1,2,.} \tag{905}$$

Autocovariances and autocorrelation function follow the same pth order difference equations as does the process itself.

Box-Jenkis (1976) approach Box-jenkins (1976) suggest three stages for implementation time series analysis - identification of lags, estimation and diagnostic checking.

- First is identification stage. Here lag order or AR(p) , MA(q) or ARMA(p, q), or ARIMA(p, d, q) is selected. It should be based on analysis of ACF and PACF functions.

Inveritibility could be examined for finite autocovariance processes. For instance the $AR(1)$ has $MA(\infty)$ representation and MA(1) has $AR(\infty)$ representation. It is better to estimate the parsimonious model.

- Second, estimation stage. Judge the goodness of fit of the model based on AIC or SBC criteria.

$$AIC = T.ln\left(\sum e^2\right) + 2n \tag{906}$$

where n is number of parameters estimated (p,q and constant); T = number of observations.

$$SBC = T.ln\left(\sum e^2\right) + n\ln(T) \tag{907}$$

4.3 Box-Jenkis (1976) approach

- Smaller the value of AIC(SBC) better is the model; for a model with perfect fit these values go to ∞. Increase the lags if the parameters fail to converge. If the series Y_t is stationary the process should converge after some iterations.
- Third important stage is diagnostic checking. Plot the residuals or standard residuals.
- Normally more than 5% values should not lie outside -2 or +2 band.
- If there is an evidence of increasing variance, it is better to use ARCH or GARCH models as the residuals from estimates are serially correlated.

Box Pierce (1970) Q statistics or Lung and Box (1978) Q statistics could be used to determine the autocorrelation of residuals.

$$BoxPierce \quad Q = T\sum r_k^2 \tag{908}$$

$$LjungBox \quad Q = T(T+2)\sum \frac{r_k^2}{T-k} \tag{909}$$

These are distributed χ_s^2. If the calculated Q statistics exceed the table χ_s^2. then the correlation coefficient is significant.

4.4 Estimation: AR(1)

Likelihood function for $AR(1)$ process

$$Y_t = c + \phi Y_{t-1} + \epsilon_t \tag{910}$$

Its mean and variances are

$$\mu = \frac{c}{1-\phi}; \quad \sigma_y^2 = \frac{\sigma^2}{1-\phi^2} \tag{911}$$

Likelihood functions:
The density of the first observation:

$$\begin{aligned} L_{y_1}(y_1, \theta) &= \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_1-\mu)^2}{\sigma_y^2}\right] \\ &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{1-\phi^2}}} \exp\left[-\frac{1}{2}\frac{\left(y_1-\frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}}\right] \end{aligned} \tag{912}$$

density for the second observation given that the first has been observed:

$$L_{y_2/y_1}(y_2/y_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2-c-\phi y_1)^2}{\sigma_y^2}\right] \tag{913}$$

$$L(y_t/y_{t-1}, y_{t-2}, ...y_{1,}\theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2-c-\phi y_{t-1})^2}{\sigma_y^2}\right] \tag{914}$$

Likelihood of the complete function

$$\begin{aligned} & L_{y_t,y_{t-1},y_{t-2},....y_1}(y_t/y_{t-1}, y_{t-2}, ...y_{1,}\theta) \\ = \; & L_{y_1}(y_1, \theta).\prod_{t=2}^{T} L(y_t/y_{t-1}, y_{t-2}, ...y_{1,}\theta) \end{aligned} \tag{915}$$

Likelihood function

$$\Im(\theta) = \log L_{y_1}(y_1, \theta) + \sum_{t=2}^{T} \log L(y_t/y_{t-1}, y_{t-2}, ...y_{1,}\theta) \tag{916}$$

Substituting the relevant functions, the likelihood for sample T Gaussian process is

$$\begin{aligned} \Im(\theta) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\phi^2}\right)\left[-\frac{1}{2}\frac{\left(y_1-\frac{c}{1-\phi}\right)^2}{\frac{\sigma^2}{1-\phi^2}}\right] + \\ & -\frac{T-1}{2}\log(2\pi) - \frac{T-1}{2} - \sum_{t=2}^{T}\left[\frac{(y_t-c-\phi y_{t-1})^2}{2\frac{\sigma^2}{1-\phi^2}}\right] \end{aligned} \tag{917}$$

4.5 Estimation: MA(1)

$MA(1)$ process

$$Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1} \tag{918}$$

$$L_{y_t}(y_{t/\epsilon_{t-1}}, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_1 - \mu - \theta\epsilon_{t-1})^2}{\sigma_y^2}\right] \tag{919}$$

$$\epsilon_1 = y_1 - \mu \tag{920}$$

$$L_{y_2/y_1}(y_2/y_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{(y_2 - c - \phi y_1)^2}{\sigma_y^2}\right] \tag{921}$$

$$\epsilon_2 = y_2 - \mu - \theta\epsilon_1 \tag{922}$$

$$\epsilon_2 = y_2 - \mu - \theta\epsilon_1 \tag{923}$$

$$\epsilon_t = y_t - \mu - \theta\epsilon_{t-1} \tag{924}$$

$$L(y_t/y_{t-1}, y_{t-2}, ...y_1, \theta) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left[-\frac{1}{2}\frac{\epsilon_t^2}{\sigma_y^2}\right] \tag{925}$$

Conditional log-likelihood for MA(1)

$$\Im(\theta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=2}^{T}\left[\frac{1}{2}\frac{\epsilon_t^2}{\frac{\sigma^2}{1-\phi^2}}\right] \tag{926}$$

Estimation: MA(p, q)
Similarly for the MA(p,q)

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + ... + \theta_p\varepsilon_{t-q} \tag{927}$$

$$\Im(\theta) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \sum_{t=2}^{T}\left[\frac{1}{2}\frac{\epsilon_t^2}{\frac{\sigma^2}{1-\phi^2}}\right] \tag{928}$$

4.5.1 Tutorial 2: Time Series, ARMA, ARIMA

Q1. Consider a monthly time series $\{y_t\}$

(a) Show how the traditional moving average based methods could be applied to decompose its trend, seasonal, cyclical and irregular components.

(b) Consider a random walk model $y_t = y_{t-1} + \varepsilon_t$ with initial conditions $y_1 = y_0$ for t =1.What are the mean, variance and the time path of y_t in terms of current and past series of errors ε_t ? What is its conditional forecast for period j made at time t? What is the error of forecast and its variance? How are the mean and variances affected if this random walk includes a drift term a_0 as in $y_t = y_{t-1} + a_0 + \varepsilon_t$.

(c) Consider signal extraction problem for series y_t including permanent and transitory shocks components as ε_t and η_t and $y_t = \varepsilon_t + \eta_t$ and $\varepsilon_t^* = a + by_t$ where $E(\varepsilon_t) = 0$; $E(\eta_t) = 0$;$E(\varepsilon_t \eta_t) = 0$; $E\left(\varepsilon_t^2\right) = \sigma^2$; $E\left(\eta_t^2\right) = \sigma^2$.
What is its minimum square error (MSE)? How is the partitioning parameters b optimally estimated?

(d) What are the prominent reasons for a failure of forecast? Illustrate Ganger and Newbold (1986) technique for combining optimal forecasts as in $f_{ct} = (1-\lambda) f_{1t} + \lambda f_{2t}$.

Q2. What is the main principle of forecasting and what are the reasons for failure of model based forecasts? Derive the forecast errors and variance of forecast for the following forecasting models .

a. Random walk with a drift: $\left[y_1 = y_0 + a_0 + \epsilon_1,\ e_{T+1} \sim N(0,1)\right]$.

b. Period h ahead forecast of AR(1): $\left[y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h},\ e_{T+h} \sim N(0,1)\right]$.

c. One period ahead forecast in MA(1): $\left[y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T,\ e_{T+1} \sim N(0,1)\right]$.

d. Two period ahead forecast in ARMA(1,1): $\left[y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1}, e_{T+2} \sim N(0,1)\right]$.

4.6 Wold Decomposition

Covariance stationary process can be decomposed into the linearly deterministic component κ_t and purely linearly indeterministic component $\sum\limits_{j=1}^{\infty} \varepsilon_{t-j}$

$$Y_t = \kappa_t + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \tag{929}$$

$$\sum_{j=1}^{\infty} \psi_j L^j = \frac{\theta(L)}{\phi(L)} = \frac{\left(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q\right)}{\left(\left(1 - \phi_1 L + \phi_2 L^2 + ... + \phi_p L^p\right)\right)} \tag{930}$$

Unit Root What is a stationary variable?

When its mean and variance are constant.

$$E(Y_t) = \mu \tag{931}$$

$$var(Y_t) = \sigma^2 \tag{932}$$

When mean and variances are not constant, that variable is non-stationary, for instance a random walk

$$Y_t = Y_{t-1} + \varepsilon_i \quad t = 1 \ldots T \tag{933}$$

In an autoregressive model

$$Y_t = \rho Y_{t-1} + \varepsilon_i \quad t = 1 \ldots T \tag{934}$$

if the autocorrelation coefficient $\rho = 1$ then it becomes a random walk. This variable is non-stationary.

A Non-Stationary variable can be made stationary by taking first difference as:

$$Y_t = \sum_{s=1}^{\infty} \rho^s \varepsilon_{t-s} \tag{935}$$

Current realisations are accumulation of past errors.

Prove that variance of this is .

$$var(Y_t) = t.\sigma^2 \tag{936}$$

Regression among non-stationary variables becomes spurious unless they are cointegrated.

$$\Delta Y_t = Y_t - Y_{t-1} \tag{937}$$

If a variable becomes stationary by taking the first difference it is said to be integrated of order one

$$I(1) \tag{938}$$

Level, drift, trend and lag terms in unit root test

If it becomes stationary after differencing d time then it is called $I(d)$ variable.

Dickey-Fuller and Phillip-Perron unit root tests are used to determine stationarity of a variable.

$$Y_t = \rho Y_{t-1} + \varepsilon_i \tag{939}$$

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \varepsilon_i; \qquad \Delta Y_t = \gamma Y_{t-1} + \varepsilon_i; \tag{940}$$

Random walk with drift

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \varepsilon_i \tag{941}$$

Level, drift, trend and lag terms in unit root test
trend stationary

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \varepsilon_i \tag{942}$$

Augmented Dickey-Fuller test

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \gamma Y_{t-1} + \sum_{i=1}^{m} \rho^s \Delta Y_{t-i} + \varepsilon_i \tag{943}$$

Cointegration in a regression

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \tag{944}$$

First do the regression and then estimate the error as

$$\widehat{\varepsilon}_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_t \tag{945}$$

Y_t and X_t are cointegrated if the estimated error is stationary $\widehat{\varepsilon}_t \sim I(0)$

$$\widehat{\varepsilon}_t = \rho \widehat{\varepsilon}_{t-1} + \varepsilon_t \tag{946}$$

if $\rho < 1$ the error $\widehat{\varepsilon}_t$ is stationary and Y_t and X_t are cointegrated. They have a long run relationship.

When variables are cointegrated there is an error correction mechanism.

$$Y_t = \varphi_2 X_t + \epsilon_t \tag{947}$$

$$Y_t = X_t + \epsilon_t \; ; \qquad \varphi_2 = 1 \tag{948}$$

4.6.1 Cointegration: Engle-Granger Representation Theorem

$$\epsilon_t = Y_t - X_t \tag{949}$$

For test of cointegration

$$\Delta \epsilon_t = \gamma \epsilon_{t-1} + u_t \tag{950}$$

$$\Delta (Y_t - X_t) = \gamma (Y_{t-1} - X_{t-1}) + u_t \tag{951}$$

$$\Delta Y_t = \Delta X_t + \gamma (Y_{t-1} - X_{t-1}) + u_t \tag{952}$$

This is an error correction model. Term $\gamma (Y_{t-1} - X_{t-1})$ gives the adjustment towards the long run equilibrium and ΔX_t denotes the short run impact.H_0 : No cointegration; t- statistics can be used instead of DF test.

4.6.2 Granger Causality Test

Estimate the following model where M_t is money Y_t is GDP and test the causality as below:

$$Y_t = \sum_{i=1}^{n} \alpha_i M_{t-i} + \sum_{j=1}^{m} \beta_j Y_{t-j} + u_{1,t} \tag{953}$$

$$M_t = \sum_{i=1}^{n} \lambda_i M_{t-i} + \sum_{j=1}^{m} \delta_j Y_{t-j} + u_{2,t} \tag{954}$$

- Unidirectional causality from M_t to Y_t requires $\sum_{i=1}^{n} \alpha_i \neq 0$ and $\sum_{j=1}^{m} \delta_j = 0$
- Unidirectional causality from Y_t to M_t requires $\sum_{j=1}^{m} \delta_j \neq 0$ and $\sum_{i=1}^{n} \alpha_i = 0$
- Bilateral causality between Y_t to M_t requires $\sum_{i=1}^{n} \alpha_i \neq 0$ and $\sum_{j=1}^{m} \delta_j \neq 0$
- Independence of Y_t to M_t from each other $\sum_{i=1}^{n} \alpha_i = 0$ and $\sum_{j=1}^{m} \delta_j = 0$

4.7 Introduction to ARCH

OLS estimates are based on the normality of errors, which are assumed to have constant mean and variance. Engel (1983) argued that many economic time series go through a series of ups and downs. Upward trend continues up to a significant length of time. and so does the downward trend.

As such the conditional mean and variance of these series are not constant. Modelling mean and vaiance of series simultaneously is the essence of the autoregressive conditional heteroskedasticity (ARCH) model.

The variance of error term is persistent and shown by autoregressive process of variances.

This technique has been widely used to measure the volatility of financial time series such as the interest rate, inflation, stock prices, returns to assets, growth rates, trends in trades.

Bollersleve (1987) modified it to generalised autoregressive conditional heteroskedasticity (GARCH) models.

ARCH test of heteroskedasticity How ARCH and GARCH models are used to test the heteroskedasticity are discussed first followed by illustrations on variants of them used to study the clustering of heteroskedastic errors commonly used in the literature.

Engle (1983) autoregressive conditional heteroskedasticy (ARCH): more useful for time series data

Model $Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \beta_3 X_{3,t} + ... + \beta_k X_{k,t} + e_t$
$\varepsilon_t \sim N\left(0, \left(\alpha_0 + \alpha_2 e_{t-1}^2\right)\right)$

$$\sigma_t^2 = \alpha_0 + \alpha_2 e_{t-1}^2 \tag{955}$$

Here σ_t^2 not observed. Simple way to estimate this is to run OLS of Y_t and get $\widehat{e}_t^2$. Then assume an ARCH (1) of errors as

$\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + v_t$ or ARCH (p) $\widehat{e}_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \alpha_3 \widehat{e}_{1-1}^2 + \alpha_4 \widehat{e}_{1-1}^2 + .. + \alpha_p \widehat{e}_{1-p}^2 + v_t$
Compute the test statistics
$n.R^2 \sim \chi_{df}^2$
Again if the calculated χ_{df}^2 is greater than table value there is an evidence of ARCH effect and heteroskedasticity.

Economies are characterised by turbulent high volatility periods followed by quite and peaceful low volatility periods.

Decision makers require some estimates of expected values as well as volatility to reflect on the uncertainties caused by such phenomenon.

Recently stock prices have risen continuously from 2002 to mid 2008 and then fell sharply in 2008 and 2009 and can be expected to rise in the next few years. Billions are lost and won because of volatilities in these series.

Engle (1987) proposes modelling expected value and volatility simultaneously by ARCH using iterative Maximum Likelihood procedure.as:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + e_t \tag{956}$$

where $e_t \sim N\left(0, \sigma_t^2\right) = N\left(0, h_t\right); \, h_t = \sigma_t^2$.

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 \tag{957}$$

A big shock in e_t in the last period have greater impact in the variance this period. The lagged effect can extend for longer periods and these are captured by putting more lagged terms on the variance equation For ARCH (2)

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 \tag{958}$$

ARCH(3)

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \alpha_3 e_{t-3}^2 \tag{959}$$

ARCH(q)

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \alpha_3 e_{t-3}^2 + ... + \alpha_q e_{t-q}^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 \tag{960}$$

4.7.1 MLE for ARCH Models

Log likelihood function for an ARCH

A likelihood function for a random error e_t is given by

$$L_t = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(\frac{-e_t^2}{\sqrt{2\sigma^2}}\right) \tag{961}$$

Joint realisations of e_1 , e_2, e_T is given by

$$L = \prod_{t=1}^{T} \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \exp\left(\frac{-e_t^2}{\sqrt{2\sigma^2}}\right) \tag{962}$$

in log form

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(e_t^2\right) \tag{963}$$

MLE for ARCH Models
where $e_t = Y_t - \beta_0 - \beta_1 X_t$.

$$\ln L = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(Y_t - \beta_0 - \beta_1 X_t\right)^2 \tag{964}$$

First order conditions for maximisations are

$$\frac{\ln L}{\partial\sigma^2} = -\frac{T}{2\sigma^2} + \frac{T}{2\sigma^4}\sum_{t=1}^{T}\left(Y_t - \beta_0 - \beta_1 X_t\right)^2 \tag{965}$$

$$\frac{\ln L}{\partial\beta_1} = \frac{T}{2\sigma^2}\sum_{t=1}^{T}\left(Y_t X_t - \beta_0 X_t - \beta_1 X_t^2\right) \tag{966}$$

For a model without intercept gives

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{T}\left(e_t^2\right)}{T} \tag{967}$$

$$\widehat{\beta}_1 = \frac{\sum_{t=1}^{T} y_t x_t}{\sum_{t=1}^{T} x_t^2} \tag{968}$$

The maximum likelihood method applied to the ARCH errors generates

$$lnL = -n\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln\left(\alpha_0 + \alpha_1 e_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{T}\frac{e_t^2}{\alpha_0 + \alpha_1 e_{t-1}^2} \tag{969}$$

MLE: ARCH

- Algorithm of ARCH process is non-linear iterative procedure. It is not possible to estimate β_0, β_1 in $Y_t = \beta_0 + \beta_1 X_{1,t} + e_t$ without knowing e_t here errors are not normal. Mean of e_t can still be zero but its variance is modelled in the variance equation.
- Thus the estimation is highly non-liner and maximum likelihood method is used to estimate this.
- First start with the initial values of $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$, and $e_t\ ..\ e_{t-S}$. Estimate $h_t = \sigma_t^2$.

- Secondly, estimate $\widehat{\beta}_0$, $\widehat{\beta}_1$ based on $\widehat{\sigma}_t^2$. Then estimate $\widehat{e}_t$ from this estimate the variance $\widehat{\sigma}_t^2$ and new values of $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$.

- Then continue the process until the values of $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\alpha}_0$, $\widehat{\alpha}_1, \widehat{\alpha}_2, ...\ \widehat{\alpha}_q$ converge.

4.7.2 GARCH tests of heteroskedasticity

Bollerslev (1987) generalised autoregressive conditional heteroskedasticy (GARCH) process is more general. For instance GARCH (1,1). Mean and variance equations take the following form:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + e_t \tag{970}$$

$$\sigma_t^2 = \alpha_0 + \alpha_2 \widehat{e}_{t-1}^2 + \beta\sigma_{t-1}^2 + v_t \tag{971}$$

GARCH (p,q) $\sigma_t^2 = \alpha_0 + \alpha_2\widehat{e}_{t-1}^2 + \alpha_3\widehat{e}_{t-2}^2 + \alpha_4\widehat{e}_{t-3}^2 + .. + \alpha_p\widehat{e}_{t-p}^2 + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2 + ..\beta_q\sigma_{t-q}^2 + .. + v_t$

Compute the test statistics $n.R^2 \sim \chi_{df}^2$

Sometimes written as

$h_t = \alpha_0 + \alpha_2\widehat{e}_{t-1}^2 + \alpha_3\widehat{e}_{t-2}^2 + \alpha_4\widehat{e}_{t-3}^2 + .. + \alpha_p\widehat{e}_{t-p}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + ..\beta_q h_{t-q} + .. + v_t$

$where \quad h_t = \sigma_t^2$

Variations of GARCH Various functional forms of h_t

$h_t = \alpha_0 + \alpha_2\widehat{e}_{t-1}^2 + \beta_1\sqrt{h_{t-1}} + v_i$ or $h_t = \alpha_0 + \alpha_2\widehat{e}_{t-1}^2 + \sqrt{\beta_1 h_{t-1} + \beta_2 h_{t-2}} + v_i$

Equivalence of GARCH(1,1) to ARCH(p)

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \delta h_{t-1} \tag{972}$$

$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \delta\left(\alpha_0 + \alpha_1 e_{t-2}^2 + \delta h_{t-2}\right)$

$$h_t = \alpha_0 + \delta\alpha_0 + \alpha_1 e_{t-1}^2 + \delta\alpha_1 e_{t-2}^2 + \delta^2 h_{t-2} \tag{973}$$

$h_t = \alpha_0 + \delta\alpha_0 + \alpha_1 e_{t-1}^2 + \delta\alpha_1 e_{t-2}^2 + \delta^2\left(\alpha_0 + \alpha_1 e_{t-3}^2 + \delta h_{t-3}\right)$

$$h_t = \alpha_0 + \delta\alpha_0 + \delta^2\alpha_0 + \alpha_1 e_{t-1}^2 + \delta\alpha_1 e_{t-2}^2 + \delta^2\alpha_1 e_{t-3}^2 + \delta^3 h_{t-3} \tag{974}$$

when the process continues $h_t = \alpha_0 + \delta\alpha_0 + \delta^2\alpha_0 + ... + \delta^j\alpha_0 + \alpha_1 e_{t-1}^2 + \delta\alpha_1 e_{t-2}^2 + \delta^2\alpha_1 e_{t-3}^2 + .. + \delta^{j-1} h_{t-3}$

$$h_t = \frac{\alpha_0}{1-\delta} + \alpha_1 \sum_{j=1}^{q} \delta^j e_{t-j}^2 \tag{975}$$

GARCH -M Model premium of holding risky asset is added to the mean equation

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \theta h_t + e_t \tag{976}$$

$e_t/\Omega_t \sim N\left(0, h_t\right)$

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \sum_{j=1}^{p} \delta^j h_t e_{t-j}^2 \tag{977}$$

Or when the mean term is in standard deviation format

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \theta\sqrt{h_t} + e_t \tag{978}$$

Threshold GARCH (TGARCH) allows to capture asymmetry between positive and negative shocks. It essentially involves introducing a dummy variable as proposed by Glosten, Jagannathan and Runkle (1993).

$$h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \theta e_{t-2}^2 d_{t-1} + \delta h_{t-1} \tag{979}$$

Here d_{t-1} is a dummy variable to isolate negative from positive shocks $d_{t-1} = 1$ is the shock is negative $e_t < 1$ and 0 otherwise. Good news and bad news have different impacts on realisation of series. Impact of good news is α_1 and of the bad news is $\alpha_1 + \theta$

Extended TGARCH

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \sum_{j=1}^{p} \delta^j h_t + \theta e_{t-2}^2 d_{t-1} \tag{980}$$

4.8 Maximum Likelihood Procedure for GARCH

Exponential GARCH (EGARCH) as proposed by Nelson (1991) models log of the variance

$$\ln(h_t) = \alpha_0 + \sum_{j=1}^{q} \psi_j \left| \frac{e_{t-j}}{\sqrt{h_{j-t}}} \right| + \sum_{j=1}^{p} \xi_j \frac{e_{t-j}}{\sqrt{h_{t-j}}} + \sum_{j=1}^{p} \delta_j h_{t-j} \tag{981}$$

It is possible to add other explanatory variables in the variance equation

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j e_{t-j}^2 + \sum_{j=1}^{p} \delta^j h_t + \sum_{j=k}^{m} \mu_k X_k \tag{982}$$

All forms of GARCH models are estimated using iterative Maximum Likelihood procedure.

$$lnL(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t = -\frac{1}{2T}\sum_{t=1}^{T} \ln(h_t) - \frac{1}{2T}\sum_{t=1}^{T} \frac{e_t^2}{h_t} \tag{983}$$

Maximum Likelihood Procedure for GARCH

Multivariate GARCH Models (based on Enders (2010))

Contemporaneous shocks to variables can be correlated with each other. Volatility spillover may occur as volatility in one variable affects that in another variable. In such situation the maximum likelihood for event t is modified as:

$$L_t = \frac{1}{2\pi\sqrt{h_{11}h_{12}(1-\rho_{12}^2)}} \exp\left[-\frac{1}{2(1-\rho_{12}^2)} \left(\frac{e_{1,t}^2}{h_{11}} + \frac{e_{1,t}^2}{h_{22}} - \frac{2\rho_{12} e_{1t} e_{1t}}{(h_{11}h_{12})^{0.5}} \right)\right] \tag{984}$$

likelihood

$$L_t = \frac{1}{2\pi \left| H \right|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2} e_t' H^{-1} e_t \right) \tag{985}$$

where $H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$
Joint density

$$L = \prod_{t=1}^{T} \frac{1}{2\pi \left| H \right|^{-\frac{1}{2}}} \exp\left(-\frac{1}{2} e_t' H^{-1} e_t \right) \tag{986}$$

$$\ln L = -\frac{T}{2} \ln\left(2\pi \right) - \frac{T}{2} \ln\left| H \right| - \frac{1}{2} \sum_{t=1}^{T} e_t' H^{-1} e_t \tag{987}$$

The maximum likelihood selects the optimal variance and covariances h_{11}, h_{12} , and h_{22} .

It is important to specify the functional forms of h_{11}, h_{12} , and h_{22}. Vech, BEK , Constant Conditional Correlation (CCC) and Dynamic Conditional Correlation are popular specifications.

4.8.1 Vech Model

$$H_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix} \tag{988}$$

$$E_t = \begin{bmatrix} e_{1,t}^2 & e_{1,t} e_{2,t} \\ e_{1,t} e_{2,t} & e_{2,t}^2 \end{bmatrix} \tag{989}$$

$$vech(\ H_t) = C + Avech(\ e_{t-1} e_{t-1}') + Bvech(\ H_{t-1}) \tag{990}$$

Diagonal vech uses only the diagonal elements and sets all $\alpha_{ij} = \beta_{ij} = 0$

4.8.2 BEK model

$$H_t = C'C + A'\ e_{t-1} e_{t-1}' A + B'\ H_{t-1} B \tag{991}$$

$$\begin{aligned} h_{11t} &= \left(c_{11}^2 + c_{12}^2 \right) + \left(\alpha_{11}^2 e_{1t-1}^2 + 2\alpha_{11}\alpha_{21} e_{1t-1} \alpha_{2t-1} + \alpha_{21}^2 e_{2t-1}^2 \right) \\ &\quad + \left(\beta_{11}^2 h_{11,t-1} + 2\beta_{11}\beta_{21} h_{12t-1} + \beta_{21}^2 h_{22,t-1} \right) \end{aligned} \tag{992}$$

$$\begin{aligned} h_{12t} &= c_{12} \left(c_{11}^2 + c_{12}^2 \right) + \alpha_{11}\alpha_{21} e_{1t-1}^2 + \left(\alpha_{11}\alpha_{22} + \alpha_{12}\alpha_{21} \right) e_{1t-1} \alpha_{2t-1} + \alpha_{21}\alpha_{22} e_{2t-1}^2 \\ &\quad + \beta_{11}\beta_{21} h_{11t-1} + (\beta_{11}\beta_{22} + \beta_{11}\beta_{21}) h_{12,t-1} + \beta_{21}\beta_{22} h_{22,t-1}) \end{aligned} \tag{993}$$

$$\begin{aligned} h_{22t} &= \left(c_{22}^2 + c_{12}^2 \right) + \left(\alpha_{12}^2 e_{1t-1}^2 + 2\alpha_{11}\alpha_{22} e_{1t-1} \alpha_{2t-1} + \alpha_{22}^2 e_{2t-1}^2 \right) \\ &\quad + \left(\beta_{12}^2 h_{11,t-1} + 2\beta_{11}\beta_{22} h_{12t-1} + \beta_{22}^2 h_{2,t-1} \right) \end{aligned} \tag{994}$$

4.8.3 CC CModel

$$H_t = \left[\begin{array}{cc} h_{11t} & \rho_{12} \left(h_{11} h_{12} \right)^{0.5} \\ \rho_{12} \left(h_{11} h_{12} \right)^{0.5} & h_{22t} \end{array} \right] \tag{995}$$

Dynamic conditional correlation (DCC) model

This is a two step procedure.

$H_t = D_t R_t D_t$

$R_t = \left(\begin{array}{cc} h_{11t}^{0.5} & 0 \\ 0 & h_{21t}^{0.5} \end{array} \right)^{-1} \left(\begin{array}{cc} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{array} \right) \left(\begin{array}{cc} h_{11t}^{0.5} & 0 \\ 0 & h_{21t}^{0.5} \end{array} \right)^{-1} = \left(\begin{array}{cc} 1 & \frac{h_{12t}}{(h_{11}h_{12})^{0.5}} \\ \frac{h_{12t}}{(h_{11}h_{12})^{0.5}} & 1 \end{array} \right)$

Now modify the likelihood function $H_t = D_t R_t D_t$

$$\ln L = -\frac{T}{2} \ln (2\pi) - \frac{T}{2} \ln |D_t R_t D_t| - \frac{1}{2} \sum_{t=1}^{T} e_t' \left(D_t R_t D_t \right)^{-1} e_t \tag{996}$$

Essentially this is equivalent to maximising $-\frac{1}{2} \sum_{t=1}^{T} 2 \ln |D_t| + e_t' D_t^{-1} D_t^{-1} e_t$ in the first stage and $-\frac{1}{2} \sum_{t=1}^{T} \ln |R_t| + e_t' \left(R_t \right)^{-1} e_t - v_t' v_t$ in the second stage where v_t are standardised residuals.

Volatility package in PcGive/ EVIEWS/RATS estimate ARCH-GARCH models.

References

[1] Asteriou D. and S. Price (2001) Political uncertainty and economic growth: UK time series evidence, Scottish Journal of Political Economy, 48:4:383-9.

[2] Bailey R W (2007)The Real Part of a Complex ARMA Process, Econometric Theory, 23, 3 , 537-545

[3] Baillie R. T. and R. J. Myers(1991) Bivariate Garch Estimation of the Optimal Commodity Futures Hedge,Journal of Applied Econometrics, 6, 2 ,109-124

[4] Bauwens L, S. Laurent and J.V. K. Rombouts (2006) Multivariate Garch Models: A Survey, Journal of Applied Econometrics, 21, 1, pp. 79-109

[5] Bollerslev T (1990) Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized Arch Model, Review of Economics and Statistics, 72, 3, 498-505

[6] Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31:307-327.

[7] Bollerslev T (1986) A conditionaly heterosckedastic time series model for speculative prices and rate of return, review of Economics and Statistics, 69, 542-547.

[8] Box G and G Jenkins (1976) Time Series Analysis, Forecasting and Conctrol, Holden Day.

[9] Breusch T S and L. G. Godfrey (1986) Data Transformation Tests, Economic Journal 96:47-58

[10] Campos J. (1986) Instrumental Variables Estimation of Dynamic Simultaneous Systems with ARMA Errors, Review of Economic Studies, 53, 1 ,125-138

[11] Dickey D.A. and W.A. Fuller (1979) Distribution of the Estimator for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, June.

[12] Dickey D.A. and W. A. Fuller (1981) Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrics, 49:4 July, 1057-1071.

[13] Engle R. (2002) New Frontiers for Arch Models Journal of Applied Econometrics, 17, 5, 425-446

[14] Engle R (2002) Dynamic conditional correlation: A simple class of multivariate Generalised autoregressive conditions heteroskedasticity model Journal of Business & Economic Statistics; Jul 2002; 20, 339-350

[15] Engle R F (2001) GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics, Journal of Economic Perspectives, 15:4:157-168

[16] Engle R F (1982) Autoregressive conditional heteroskedasticity with estimates of variance of UK Inflation, Econometrica, 50:3:987-1008.

[17] Engle R F (1983) Estimates of the Variance of U. S. Inflation Based upon the ARCH Model, Journal of Money Credit and Banking, 15:3:286-301.

[18] Engle R E and C.W.J. Granger (1987) Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica, 55, 2, 251-276.

[19] Engle R. F. and K. F. Kroner (1995) Multivariate simultaneous Generalized ArchEconometric Theory, 11, 1995, 122-150.

[20] Engel R F, D. F. Hendry and D Trumple (1985) Small-Sample Properties of ARCH Estimators and Tests, Canadian Journal of Economics, 18:1:66-93.

[21] Enders W. (2010) Applied Econometric Time Series, 3rd ed., Willey.

[22] Evans G and S Honkapohja (1986) A Complete Characterization of ARMA Solutions to Linear Rational Expectations Models,Review of Economic Studies, 53, 2, 227-239

[23] Hamilton James D. (1994) Time Series Analysis, Princeton.

[24] Harris R and R Sollis (2003) Applied Time Series Modelling and Forecasting, John Wiley and Sons.

[25] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[26] Harvey A. edt. (1994) Time Seris Analysis, Edward Elgar.

[27] Harvey A.C. (1989) Forecasting structural time series models and Kalman Filter, Cambridge University Press.

[28] Harvey A. (1976) Estimating Regression Models with Multiplicative Heteroskedasticity, Econometrica, 44:461-465.

[29] Huang D, H. Wang and Q Yao (2008) Estimating GARCH models: when to use what? Econometrics Journal, 11, 27–38

[30] Kleibergen F. and H. K. Van Dijk(1993) Non-Stationarity in GARCH Models: A Bayesian Analysis Journal of Applied Econometrics, 8, pp. S41-S61

[31] Koopman S. J., J. A. Doornik, A. C. Harvey and N Shephard (2012) STAMP 8.2, Timberlake

[32] Ling S. and Michael McAleer (2003) Asymptotic Theory for a Vector ARMA-GARCH Model, Econometric Theory, 19, 2 , 280-310

[33] Lütkepohl H (1984) Forecasting Contemporaneously Aggregated Vector ARMA Processes,Journal of Business & Economic Statistics, 2, 3, 201-214

[34] Lumsdaine R. L (1995) Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1) Models: A Monte Carlo Investigation,Journal of Business & Economic Statistics, 13, 1,1-10

[35] McCabe B. P. M. and S. J. Leybourne (1998) On Estimating an ARMA Model with an MA Unit Root, Econometric Theory, 14, 3, 326-338

[36] Mills T.C. (2003) Modelling Trends and Cycles in Economic Time Series, Palgrave.

[37] Mills T.C. (2000) The Econometric Modelling of Financial Time Series, Cambridge University Press.

[38] Nelson (1991) Conditional heteroskedasticity in asset returns: a new approach, Econom

[39] Nelson C. R. and C. I. Plosser (1982) Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics.

[40] Pagan A. and M. Wickens (1989) A Survey of Some Recent Econometric Methods, Economic Journal, 99 962-1025.

[41] Pan J, Hui Wang and Qiwei Yao (2007) Weighted Least Absolute Deviations Estimation for ARMA Models with Infinite Variance Econometric Theory, 23, 5, 852-879

[42] Pesaran MH, Shin Y, Smith RJ.(2001), Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, 16, 289-326

[43] Phillips P.C.B. (1987) Time Series Regression with an Unit Root, Econometrica, 55, 2, 277-301.

[44] Prescott, E.C. (1986), "Theory Ahead of Business Cycle Measurement," Federal Reserve Bank of Minneapolis, Quarterly Review; Fall.

[45] Pyndick RS and DL Rubinfeld (1998) Econometric Models and Economic Forecasts, McGraw Hill

[46] Sims, C. A. and Uhlig, H. (1991) Understanding unit rooters: a helicopter tour, Econometrica, 59, 1591-9.

[47] Stock J. H. and M. Watson (2002) Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association, 97, 1167-1179.

[48] Wooldridge J M (1994) Efficient Estimation Under Heteroskedasticity, Econometric Theory, 10:1:223

[49] White H (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direst Test for Heteroskedasticity, Econometrica 48:4:817-838.

[50] http://www.feweb.vu.nl/econometriclinks/

4.9 Interest Rate Determination Rule: Taylor Rule

Output gap and interest rate

$$(y_t - y_t^*) = -d\,(i_{t-1} - i^*)\,; \quad d > 0 \tag{997}$$

Inflation and output (Supply or Phillips curve)

$$(\pi_t - \pi_t^*) = c\left(y_{t-1} - y_{t-1}^*\right); \quad c > 0 \tag{998}$$

Interest rate determination rule

$$i_t = i^* + a\,(y_t - y_t^*) + b\,(\pi_t - \pi_t^*)\,; \quad a > 0,\ b > 0 \tag{999}$$

Solution of the Interest Rate Rule Model

$$\begin{aligned} i_t &= i^* + a\,(y_t - y_t^*) + b\,(\pi_t - \pi_t^*) \\ &= i^* - ad\,(i_{t-1} - i^*) + cb\left(y_{t-1} - y_{t-1}^*\right) \\ &= i^* - ad\,(i_{t-1} - i^*) - cbd\,(i_{t-2} - i^*) \end{aligned} \tag{1000}$$

Collecting terms

$$i_t + ad.i_{t-1} + cbd.i_{t-2} = i^* + ad.i^* + cbd.i^* \tag{1001}$$

Iterating forward by two periods

$$i_{t+2} + ad.i_{t+1} + cbd.i_t = i^* + adi^* + cbdi^* \tag{1002}$$

Long run natural rate of interest: steady state

$$i_t = i_{t-1} = i_{t-2} = \widehat{i} \tag{1003}$$

$$(1 + ad + cbd)\widehat{i} = i^*(1 + ad + cbd) \tag{1004}$$

$$\widehat{i} = i^* \tag{1005}$$

Fluctuations around this long run interest rate depends on homogenous part of the second order difference equation

$$i_{t+2} + ad.i_{t+1} + cbd.i_t = 0 \tag{1006}$$

Transitional dynamics (replace $i_t = Ab^t$ in homogenous equation).

$$Ab^{t+2} + ad.Ab^{t+1} + cbd.Ab^t = 0 \tag{1007}$$

$$b^2 + ad.b + cbd = 0 \tag{1008}$$

Three Cases in time path of the interest rate model
Cycle depends on roots of the quadratic equation

$$b_1, b_2 = \frac{-ad \pm \sqrt{(ad)^2 - 4cbd}}{2} \tag{1009}$$

Distinct real root case (no cycle)

$$(ad)^2 > 4cbd \tag{1010}$$

Repeated real root case (no cycle)

$$(ad)^2 = 4cbd \tag{1011}$$

Complex root case (cycle)

$$(ad)^2 < 4cbd \tag{1012}$$

Complete solution

$$i_t = A_1 b_1^t + A_2 b_2^t + \widehat{i} \tag{1013}$$

Example of Complex Root Case: Example
Preliminaries $i_t = A_1 R^t (\cos\theta \cdot t + i \cdot \sin\theta \cdot t) + A_2 R^t (\cos\theta \cdot t - i \cdot \sin\theta \cdot t) + \widehat{i}$
Exponential forms and polar coordinates

$$R = \sqrt{h^2 + v^2} = bcd \tag{1014}$$

$$\sin\theta = \frac{v}{R} \implies v = R.sin\theta \tag{1015}$$

$$\cos\theta = \frac{h}{R} \implies h = R.co\theta \tag{1016}$$

$$e^{i\theta} = \cos\theta + i\,\mathrm{Si}\,n\theta \qquad e^{-i\theta} = \cos\theta - i\,\mathrm{Si}\,n\theta \tag{1017}$$

$$h \pm vi = R.co\theta \pm R.i\sin\theta = R.(co\theta \pm i\sin\theta) = \mathrm{Re}^{\pm i\theta} \tag{1018}$$

$\frac{\partial \sin\theta}{\partial\theta} = \cos\theta$; $\frac{\partial \cos\theta}{\partial\theta} = -\sin\theta$;
Multiplier Accelerator Model: Complex Root
$(ad)^2 < 4cbd$

Need to consider the algebra for the imaginary number and some trigonometric functions in this case. Using Pythagorean in an imaginary axis is used to derive the roots of the characteristic equation.

$$b_1, b_2 = (h \pm v \cdot i) = -\frac{ad}{2} \pm i\sqrt{\frac{4cbd - (ad)^2}{2}} \tag{1019}$$

$$Y_t = A_1 b_1^t + A_2 b_2^t = A_1 (h + v \cdot i)^t + A_2 (h - v \cdot i)^t \tag{1020}$$

Using DeMoivre's theorem

$$(h \pm v \cdot i) = R^{ht} (\cos\theta \cdot t \pm i \sin\theta \cdot t) \quad \text{for} \quad R^{ht} > 0. \tag{1021}$$

Multiplier Accelerator Model: Complex Root
Imaginary axis (Pithagorus Theorem)

$$R = \sqrt{h^2 + v^2} = \alpha\gamma \tag{1022}$$

$$i_t = A_1 R^{ht} (\cos\theta \cdot t + i \sin\theta \cdot t) + A_2 R^{ht} (\cos\theta \cdot t - i \sin\theta \cdot t) \tag{1023}$$

$$i_t = A_1 R^{ht} \left(\cos\frac{\pi}{2} \cdot t + i \sin\frac{\pi}{2} \cdot t\right) + A_2 R^{ht} \left(\cos\frac{\pi}{2} \cdot t - i \sin\frac{\pi}{2} \cdot t\right) \tag{1024}$$

Three possibilities:
i) $R^{ht} > 1$; $\alpha\gamma > 1$ ii) $R^{ht} = 1 \quad \alpha\gamma = 1$ and ii) $R^{ht} < 1$ $\alpha\gamma < 1$ Only the $\alpha\gamma < 1$ case is convergent other two cases are divergent.

5 L4: Cointegration

Cointegration is the long run relationship among variables. Engle and Granger (1987) are attributed with this idea but the concepts evolved over time with works of Dickey and Fuller (1979), Phillips (1987), Johansen(1988), Harvey (1990), Kuthbertson, Hall andTaylor (1992), Banerjee , Dolado and Galbraith and Hendry (1993), Hamilton (1994) Harris (1995), Patterson (2000) Harris and Sollis (2003), Doornik and Hendry ((2003), Greene (2008). Enders(2010). Thus the evolution of the literature in cointegration analysis could be stated as:

- Engle-Granger (1987), Johansen (1988), Phillips and Ouliaris (1990), Park (1992), Hansen (1992), Hendry and Doornik (1994)

- Nelson and Plosser (1982), Campbell and Shiller (1987), Enders (1988), Hylleberg and Mizon (1989), von Hagen (1989)

- Kim (1990), Davidson and Hall (1991), Hall, Anderson and Granger (1992), Solo (1995), Balke and Fomby (1997)

- Enders and Siklos (2001), Villani (2005), Hualde (2006), Qu and Perron (2007), Westerlund (2008), Bhattarai (2008)

- Jawadi, Bruneau and Sghaier (2009), Turner (2009), Apergis, Dincer and Payne (2010), Trapani and Urga (2010), Bansal and Kiku (2011)

- Texts: Harvey (1990), Kuthbertson, Hall and Taylor (1992), Banerjee, Dolado, Galbraith and Hendry (1993), Hamilton (1994) Harris and Sollis (2003), Enders (2010).

Let us cosider a simple interest rate rule model that can be estimated following Johansen and Juselius (1990) procedure for a cointegrated VAR model. The validity of this approach is based on the rank of the cointegration matrix of the structural coefficients that is crucial for determining the number of cointegration vectors in the model. Consider a VAR model for above three variables.

$$Y_t = AY_{t-1} + e_t \tag{1025}$$

where Y_t is the vector of interest rate, output gap and inflation gap and e_t is the vector of normally and identically distributed random error terms. By subtracting Y_{t-1} from both sides

$$\Delta Y_t = (A - I)\, Y_{t-1} + e_t = \Pi Y_{t-1} + e_t \tag{1026}$$

where $\Pi = (A - I)$.

Here Π is the matrix of parameters showing the total long run relationship among variables. By using the cointegration procedure this matrix can further be decomposed into adjustment coefficients (α) and cointegrating vectors (β) as $\Pi = \alpha\beta$. The β matrix denotes the long run steady state relationship among variables and α is the dynamic process of adjustment towards that equilibrium. The estimation on interest rate, output gap and inflation gap for the UK for 1972:2 to 1999:4 obtained in Bhattarai (2008) using the PcGive (Doornik and Hendry (2001)) yields following results.

$$\Pi = \begin{bmatrix} -0.037 & -0.120 & 0.186 \\ 0.029 & -0.242 & -0.085 \\ -0.094 & -0.072 & -0.537 \end{bmatrix}$$

$$\alpha = \begin{bmatrix} 0.018 & 0.091 & 0.011 \\ 0.017 & -0.003 & -0.002 \\ -0.007 & -0.210 & 0.007 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 1.000 & 0.195 & -6.646 \\ -13.850 & 1.000 & 3.636 \\ -4.368 & 2.790 & 1.000 \end{bmatrix}$$

The number of co-integrating vectors in the Johansen procedure is determined by $\lambda_{trace(r)} = -T \sum_{i=r+1}^{n} \ln\left(1 - \widehat{\lambda}_i\right)$ and $\lambda_{\max(r,r+1)} = -T \ln\left(1 - \widehat{\lambda}_{r+1}\right)$ statistics, where $\widehat{\lambda}_i$ denotes the eigenvalues of the characteristic matrix $\Pi = (A - I)$ and r is an indicator for a reduced rank in (k-r) for k number of explanatory variables. The calculated values of these statistics are compared with the theoretical critical values from Johansen and Juselius (1990) to ascertain the number of cointegrating ranks as following.

These cointegration results are comparable to those found in other applied works such as Cheung and Westerman (2002), Yamada (2002), Brooks and Skinner (2000),Camarero, Ordonez and Tamarit (2002) and Silvapulle and Hewarathna (2002), Valente (2003), Mills and Wood (2002).

Table 18: Cointegration test results

Rank: H_0	Trace test [Prob]	Max test [Prob]	Tracetest (T-nm)	Max test (T-nm)
r = 0	56.86 [0.000]**	34.38 [0.000]**	55.43 [0.000]**	33.52 [0.000]**
r $\leq$ 1	22.48 [0.003]**	12.68 [0.087]	21.91 [0.004]**	12.36 [0.097]
r $\leq$ 2	9.80 [0.002]**	9.80 [0.002]**	9.55 [0.002]**	9.55 [0.002]**

The order of the rank of Π suggests the number of cointegrating vectors in β. Above $\lambda_{trace(r)}$ and $\lambda_{\max(r,r+1)}$ tests suggest that at least there are two cointegrating vectors in the above model. The long run relation among these variables is shown by a very good fit of the predicted and actual series of above three variables.

This section takes on technical issues involved in these ideas systematically.

5.0.1 Order of integration

- Variables may trend up and down but they may move together so that they have some linear relationship or Cointegration .
- There are mainly two ways to study cointegration among variables. Two step procedure adopted by Engle-Granger (1987) useful mainly for single equation models and Johansen (1988) procedure for multiple equation models.
- Consider X a vector of variables. If each variable has the order of integration equal to d then $X_t \sim I(d)$; it will be stationary after differencing d times. Then consider a linear combinations $Z_t = \beta X_t \sim I(b)$ for any $b > 0$. The linear combination X variables denoted by β make it to integrate by the order of $(d - b)$ then the β is a cointegrating vector.

Cointegration: Definition

- If a normal regression generates a stationary white noise $e_t \sim N(0, \sigma^2)$ as below then β : $(\beta_1, \beta_2, ..., \beta_n)$ are cointegrating vector.

$$\beta_1 x_{1,t} + \beta_2 x_{2,t} + ... + \beta_n x_{n,t} = e_t \tag{1027}$$

- One popular example illustrated in many texts is demand for money equation:

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t \tag{1028}$$

- Assume that each variable included in the equation m_t p_t, y_t, r_t is $I(1)$ but their linear combination is $I(0)$ as expressed below.

$$e_t = m_t - \beta_0 - \beta_1 p_t - \beta_2 y_t - \beta_3 r_t \quad \sim I(0) \tag{1029}$$

- This implies the these variables have long run relationship and they are cointegrated. The cointegration vector is given by

$$\beta : [\beta_0, \beta_1, \beta_2, \beta_3] \tag{1030}$$

Integration and Cointegration

- Thus cointegration in Engle Granger(1987) terminology is linear combination of nonstationary variables indicating long-run equilibrium relationship that may reflect casual, behavioral or simply reduced form relation.

- Consider three variables with different order of integration $X_t \sim I(d)$, $Y_t \sim I(e)$ for $e > d$ .and a resulting series which is the linear combinations of these two $z_t = \alpha X_t + \beta Y_t$ will integrate by higher order of two.

$$\Delta^d z_t = \alpha \Delta^d X_t + \beta \Delta^d Y_t \tag{1031}$$

- Here $\Delta^d X_t$ will be stationary but $\Delta^d Y_t$ still non-stationary. Therefore $\Delta^d z_t$ is also non-stationary; requires further differencing.

$$\Delta^e z_t = \alpha \Delta^e X_t + \beta \Delta^e Y_t \tag{1032}$$

Now both the LHS and the RHS terms are stationary. Thus $z_t \sim I(\max(d, e))$.

- Thus even if the series are trended they move closely so that the differences between them is constant (stationary).

- Error terms between them have well defined first and second moments and the OLS is feasible even among non-stationary variables if they are cointegrated.

- Such cointegration reflects long term equilibrium relationship. components of two or more series exactly offset each other .

- To sum up components of vector X_t are said to be cointegrated $X_t \sim CI(d, b)$ if i) $X_t \sim I(d)$ and ii) $Z_t \sim \alpha' X_t \sim I(d - b)$ for $b > 0$

- In other words if a set of $I(d)$ variables X_t yields a linear combination αX_t that has a lower order of integration $(d - b) < d$ for $b > 0$ then vector α is a cointegrating vector.

Integration and Cointegration

- Series integrated of different order cannot be cointegrated: if $X_t \sim I(0)$, $Y_t \sim I(1)$ the resulting series $Y_t - \alpha X_t$ will still have a drift and becomes infinitely large.

- Introducing new instruments could solve this problem.$Y_t \sim I(1)$, $X_t \sim I(2)$, $W_t \sim I(2)$ here $X_t \sim I(2)$ and $W_t \sim I(2)$ cointegrate as

$$V_t = \alpha X_t + c W_t \sim I(I) \tag{1033}$$

- Then it is possible to have $Z_t = eV_t + fY_t \sim I(0)$ then $X_t, W_t \sim CI(2, 1)$, $V_t, Y_t \sim CI(1, 1)$ and $Z_t \sim I(0)$.

5.0.2 Engle-Granger Representation Theorem (EGRT) and Error Correction Model

If a series of variables are cointegrated of order 1, $CI(1,1)$ then there exists a valid error correction representation of the data.

Let X_t be $N \times 1$ vector with $I(1)$ then $\alpha X_t \sim I(0)$ then the error correction representation is given by:

$$\Phi L\,(1-L)\,X_t = -\alpha' X_{t-1} + \Theta L\,(e_t) \tag{1034}$$

Proof of EGRT

$$\Delta X_t = \Pi_0 + \Pi_1 X_{t-1} + \Pi_1 \Delta X_{t-1} + \Pi_2 \Delta X_{t-2} + ... + \Pi_p \Delta X_{t-p} + \varepsilon_t \tag{1035}$$

Rewrite this as:

$$\Pi X_{t-1} = \Delta X_t - \Pi_0 - \Pi_1 \Delta X_{t-1} - \Pi_2 \Delta X_{t-2} - .... - \Pi_p \Delta X_{t-p} - \varepsilon_t \tag{1036}$$

X_t are $I(1)$, therefore ΔX_{t-j} are $I(0)$. In the above equation the right hand side (RHS) variables are stationary.

Engle-Granger Representation Theorem (EGRT)and Error Correction

In order to maintain the time series properties, LHS should also be stationary. X_t are $I(1)$ therefore $\Pi : [\Pi_0, \Pi_1, \Pi_2, ...\Pi_p]$ must be cointegrating vector. By a bit expansion this means

$$\Pi_{11} X_{1,t-1} + \Pi_{12} X_{2,t-1} + \Pi_{13} X_{t-3} + ... + \Pi_{1,p} X_{t-p} \quad \sim \quad I(0) \tag{1037}$$

It is inappropriate to estimate the VAR of cointegrated variables only in the first differences, it also needs level terms ΠX_{t-1} for error correction.

Error Correction Model Consider a stylized model of error correction:

$$Y_t = \varphi_2 X_t + \epsilon_t \tag{1038}$$

$$Y_t = X_t + \epsilon_t \; ; \qquad \varphi_2 = 1 \tag{1039}$$

$$\epsilon_t \;= Y_t - X_t \tag{1040}$$

For test of cointegration

$$\Delta \epsilon_t \;= \gamma \epsilon_{t-1} \; + u_t \tag{1041}$$

$$\Delta\,(Y_t - X_t) = \gamma\,(Y_{t-1} - X_{t-1}) + u_t \tag{1042}$$

$$\Delta Y_t = \Delta X_t + \gamma\,(Y_{t-1} - X_{t-1}) + u_t \tag{1043}$$

This is an error correction model.

- H_0 : No cointegration; t- statistics can be used instead of DF test.
- Term $\gamma\,(Y_{t-1} - X_{t-1})$ gives the adjustment towards the long run equilibrium
- and ΔX_t denotes the short run impact.

5.0.3 A simple example of demand for oil

OLS _ECM for long run elasticities wrt price and income

$$LnD_{i.t} = \alpha_i + \beta_i LnP_{i.t} + \gamma_i LnY_t + u_{it} \tag{1044}$$

Error correction mechanism and short run elasticities:

$$\Delta LnD_{i.t} = \delta_i + \theta_i \Delta LnP_{i.t} + \lambda_i \Delta LnY_t + \rho_i ECM_{i,t} + e_{it} \tag{1045}$$

For forecasting differentiae (1044) wrt time

$$\widehat{D}_{i.t} = \beta_i \widehat{P}_{i.t} + \gamma_i \widehat{Y}_t \Longrightarrow D_{i.t+1} = D_{i.t} + \widehat{D}_{i.t} = D_{i.t} + \beta_i \widehat{P}_{i.t} + \gamma_i \widehat{Y}_t \tag{1046}$$

These are very standard in the literature.
Agrawal's empirical results on demand elasticities for India

Table 19: Estimated price and income elasticities

	1970-2011		1970-2006	
D	β_i	γ_i	β_i	γ_i
Crude oil	-0.41	1.00	-0.34	1.00
Diesel	-0.56	1.02	-0.58	1.01
Petrol	-0.85	1.39	-0.82	1.38

- Literature on energy
  - India: Kumar and Jain (2010), Sajal Ghosh (2006, 2009, 2010), Goldar and Mukhopadhyay (1990), Ramanathan (1999), Rao and Parikh (1996), Parikh,Purohit, and Maitra (2007).
  - Adams and Shachmurove (2008) for China; Altinay (2007) for Turkey.
  - cointegration: Johansen and Juselius (1990), Pesaran, Shin and Smith (2001)
- Further literature on energy
  - Hamilton's classic model on energy and growth; RICE and DICE models by Nordhaus; IEEA studies
  - IAEE sessions on Jan 4th (Sieminski, El-Gamal; Killian; Parsons)

5.0.4 LSE Tradition on Dynamic Modelling

LSE Tradition in Dynamic Modelling (as discussed in Cuthbertson, Hall and Taylor(1992))

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \tag{1047}$$

Restrictions and implications
1) $\alpha_1 = 0$; $\beta_1 = 0$ static regression model
2) $\beta_1 = 0$ partial adjustment model

3) $\alpha_1 = 0$; $\beta_0 = 0$ X is a leading indicator of Y
4) $\alpha_1 = 0$ finitely distributed lag model
5) $\alpha_1 = 1$; $\beta_0 + \beta_1 = 0$ first difference formulation
6) $\beta_0 = 0$ Y as ARDL(1,1)
Taking lags

$$(1 - \alpha_1 L) Y_t = \alpha_0 + \beta_0 \left(1 + \frac{\beta_1}{\beta_0} L\right) X_t + u_t \tag{1048}$$

with restriction $\alpha_1 = -\frac{\beta_1}{\beta_0}$and when $(L = 1)$

$$Y_t = \alpha^* + \beta_0 X_t + e_t \tag{1049}$$

where $\alpha^* = \frac{\alpha_0}{(1-\alpha_1)}$; $e_t = (1 - \alpha_1 L)^{-1}$.

The Error correction representation of this model is obtained by subtracting Y_{t-1} both sides, adding and subtracting $\beta_0 X_{t-1}$,$\alpha_1 X_{t-1}$ in the RHS and rearranging the equation

$$\Delta Y_t = \alpha_0 + \beta_0 \Delta X_t - (1 - \alpha_1)(Y_{t-1} - X_{t-1}) + \gamma X_{t-1} + u_t \tag{1050}$$

where $\gamma = \alpha_1 + \beta_0 + \beta_1 - 1$

Both (1047) and (1050) are different ways of expressing the same relation but the (1050) is more in the format of the error correction model. For static solution there are no changes in the variables $\Delta X_{t-j} = \Delta Y_{t-j} = 0 \;\; X_{t-j} = X$ and $Y_{t-j} = Y$

$$Y = \frac{\alpha_0}{(1 - \alpha_1)} + \frac{\beta_0 + \beta_1}{(1 - \alpha_1)} X \tag{1051}$$

Further when $\gamma = 0 \Rightarrow \beta_0 + \beta_1 = (1 - \alpha_1)$

$$Y = \frac{\alpha_0}{(1 - \alpha_1)} + X \tag{1052}$$

Then the dynamic equation is

$$\Delta Y_t = \beta_0 \Delta X_t - (1 - \alpha_1)\left(Y_{t-1} - Y^*_{t-1}\right) \tag{1053}$$

where $Y^*_{t-1} = \frac{\alpha_0}{(1-\alpha_1)} + X_{t-1}$ which is the long run equilibrium value.

If $(1 - \alpha_1) > 0$ and $Y_{t-1} > Y^*_{t-1}$ then ΔY_t falls in the next period. It brings Y_t closer to Y^*_{t-1}. Estimated dynamic equation

$$\Delta Y_t = \alpha_0 + \beta_0 \Delta X_t + \beta_1 (Y_{t-1} - X_{t-1}) \tag{1054}$$

whereβ_1=$-(1 - \alpha_1)$

Thus agents adjust their behavior according to singnals that they are out of equilibrium. For instance if $(Y_{t-1} - X_{t-1})$ is the ratio of stock of money to the income, deviation of money income ratio from the equilibrium will lead to future changes in money holding by agent in order to move closer to the desired long run equilibrium.

Example 2

$$Y_t = \alpha_0 + \gamma_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \tag{1055}$$

Long-run

$$Y = \frac{\alpha_0}{(1-\gamma_1)} + \frac{\beta_0+\beta_1}{(1-\gamma_1)}X + \frac{\epsilon_t}{(1-\gamma_1)} \tag{1056}$$

$$Y = \varphi_0 + \varphi_1 X + \varepsilon_t \tag{1057}$$

$\varphi_0 = \frac{\alpha_0}{(1-\gamma_1)}$; $\varphi_1 = \frac{\beta_0+\beta_1}{(1-\gamma_1)}$; $\varepsilon_t \sim N(0,\sigma^2)$

To reparametise this in the error correction form first subtract Y_{t-1} from both sides and add and subtract $\beta_0 X_{t-1}$ and $\gamma_1 X_{t-1}$from the right hand side

$$\begin{aligned} Y_t - Y_{t-1} &= \alpha_0 + \gamma_1 Y_{t-1} - Y_{t-1} + \beta_0 X_t + \beta_0 X_{t-1} - \beta_0 X_{t-1} \\ &\quad + \beta_1 X_{t-1} + \gamma_1 X_{t-1} - \gamma_1 X_{t-1} + \epsilon_t \end{aligned} \tag{1058}$$

$\Delta Y_t = \alpha_0 - (1-\gamma_1) Y_{t-1} + \beta_0 \Delta X_t + \beta_0 X_{t-1} - X_{t-1} + \beta_1 X_{t-1} + \gamma_1 X_{t-1} + X_{t-1} - \gamma_1 X_{t-1} + \epsilon_t$
$\Delta Y_t = \alpha_0 - (1-\gamma_1) Y_{t-1} + \beta_0 \Delta X_t + \beta_0 X_{t-1} - X_{t-1} + \beta_1 X_{t-1} + \gamma_1 X_{t-1} + (1-\gamma_1) X_{t-1} + \epsilon_t$
$\Delta Y_t = \alpha_0 - (1-\gamma_1)(Y_{t-1} - X_{t-1}) + \beta_0 \Delta X_t + \theta_0 X_{t-1} + \epsilon_t$
where $\theta_0 = \beta_0 + \beta_1 + \gamma_1$.
Uses of Error Correction Mechanism
$\Delta Y_t = \beta_0 \Delta X_t - \beta_1 (Y_{t-1} - 0.9 X_{t-1})$
Long run $Y_{t-1} = 0.9 X_{t-1}$

5.0.5 Three ways of error correction

Three ways of representing error correction mechanism
ADL(1,1)

$$Y_t = \alpha_0 + \gamma_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \tag{1059}$$

ECM: non-linear in coefficients:

$$\Delta Y_t = \alpha_0 - \left(1-\gamma_1\right)(Y_{t-1} - X_{t-1}) + \beta_0 \Delta X_t + \theta_0 X_{t-1} + \epsilon_t \tag{1060}$$

Bewley Transformation

$$Y_t = \gamma_1 + \gamma_2 X_t + \beta_1^* \Delta X_t + \gamma_1^* \Delta Y_{t-1} + \epsilon_t^* \tag{1061}$$

Equilibrium Errors
Proof:
$Y_t - \gamma_1 Y_{t-1} = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$
$\beta_0 X_t + \beta_1 X_{t-1} = (\beta_0 + \beta_1) X_t - \beta_1 (X_t - X_{t-1}) = (\beta_0 + \beta_1) X_t - \beta_1 \Delta X_t$
$Y_t - \gamma_1 Y_{t-1} = \left(1-\gamma_1\right) Y_t - \gamma_1 (Y_t - Y_{t-1}) = \left(1-\gamma_1\right) Y_t - \gamma_1 \Delta Y_t$
Substituting these in the parent function
$Y_t = \frac{\alpha_0}{(1-\gamma_1)} + \frac{\beta_0+\beta_1}{(1-\gamma_1)} X_t - \frac{\beta_1}{(1-\gamma_1)} \Delta X_t - \frac{\gamma_1}{(1-\gamma_1)} \Delta Y_t + \frac{\epsilon_t}{(1-\gamma_1)}$
$\varphi_1 = \frac{\alpha_0}{(1-\gamma_1)}$; $\varphi_2 = \frac{\beta_0+\beta_1}{(1-\gamma_1)}$; $\beta_1^* = -\frac{\beta_1}{(1-\gamma_1)}$; $\gamma_1^* = -\frac{\gamma_1}{(1-\gamma_1)}$; $\epsilon_t^* = \frac{\epsilon_t}{(1-\gamma_1)}$;

$$Y_t = \varphi_1 + \varphi_2 X_{t-1} + \beta_1^* \Delta X_t + \gamma_1^* \Delta Y_t + \epsilon_t^* \tag{1062}$$

Equilibrium errors is thus

$$\epsilon_t^* = Y_t - (\varphi_1 + \varphi_2 X_{t-1} + \beta_1^* \Delta X_t + \gamma_1^* \Delta Y_t) \tag{1063}$$

5.0.6 Johansen cointegration analysis

Start with X_t . Let it be vector of $N \times N$ dimension each integrated of order 1 $I(1)$. The VAR is given by

$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + ... + \Pi_k \Delta X_{t-k} + \varepsilon_t$

Π_i is $N \times K$ matrix of coefficients This implies

$$\begin{pmatrix} X_{1,t} \\ . \\ . \\ X_{n,t} \end{pmatrix} = \begin{pmatrix} \Pi_{111} & \Pi_{112} & . & . & . & \Pi_{11N} \\ \Pi_{121} & . & . & . & . & \Pi_{12N} \\ . & . & . & . & . & . \\ \Pi_{1N1} & \Pi_{N21} & . & . & . & \Pi_{1NN} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ . \\ . \\ X_{n,t-1} \end{pmatrix} + ...$$

$$.... + \begin{pmatrix} \Pi_{11k} & \Pi_{12k} & . & . & . & \Pi_{1Nk} \\ \Pi_{21k} & . & . & . & . & \Pi_{21k} \\ . & . & . & . & . & . \\ \Pi_{N1k} & \Pi_{N2k} & . & . & . & \Pi_{NNk} \end{pmatrix} \begin{pmatrix} X_{1,t-k} \\ . \\ . \\ X_{n,t-k} \end{pmatrix}$$

Johansen Cointegration Analysis

ECM representation of this VAR is given by

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1064}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_1 + ... + \Pi_i \quad \text{for } i = 1..k$

Γ_k gives the long run solution.

- Here the terms in LHS and the first $k-1$ elements of difference of X_t are $I(0)$. The last element $\Gamma_k X_{t-k}$ is the linear combination of $I(1)$ variables it must be stationary to balance the time series properties of right and left hand side.
- Johansen uses the canonical correlation method to produce all the distinct combinations of level of X which produce high correlation with $I(0)$ elements in above equations. These correlates are cointegrating vectors.
- It is very instructive to develop intuition of cointegration method following examples provided by Patterson (2000).

Cointegration Analysis: Example 1 A simple ECM without exogenous X variables with $Y_{1,t}$ and $Y_{2,t}$ as $I(1)$

$$\Delta Y_{1,t} = -\frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8} Y_{2,t-1}\right) + \epsilon_{1,t} \tag{1065}$$

$$\Delta Y_{2,t} = \frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8} Y_{2,t-1}\right) + \epsilon_{2,t} \tag{1066}$$

$\Delta Y_{1,t} = \Delta Y_{2,t} = 0$ in the steady state; $E(\epsilon_{1,t}) = 0$; $E(\epsilon_{2,t}) = 0$. This implies the equilibrium relations $Y_{1,t-1} - \frac{1}{8} Y_{2,t-1} = 0$ and the equilibrium error is given by $\xi_t = Y_{1,t-1} - \frac{1}{8} Y_{2,t-1}$. if $\xi_t \neq 0$ there is a disequilibrium in the last period causing $\Delta Y_{1,t}$ $\Delta Y_{2,t}$ to change to correct the equilibrium.

The adjustment coefficient for $\Delta Y_{1,t}$ is $-\frac{1}{2}$. That means if $\xi_t > 0 \Rightarrow Y_{1,t-1} > \frac{1}{8}Y_{2,t-1}$ so $\Delta Y_{1,t}$ should decrease to move towards equilibrium. Conversely if $\xi_t < 0 \Rightarrow Y_{1,t-1} < \frac{1}{8}Y_{2,t-1}$ so $\Delta Y_{1,t}$ should increase to move toward equilibrium.

The adjustment coefficient in $\Delta Y_{2,t}$ is positive $\frac{1}{2}$. if $\xi_t > 0 \Rightarrow \Delta Y_{2,t-1} > 0$. It moves closer to the equilibrium.

With two variables only one equilibrium relation can be defined though it can be parameterised in many ways.

$\frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) = -\frac{1}{16}\left(Y_{2,t-1} - 8Y_{1,t-1}\right)$

The second equation could have been written as

$$\Delta Y_{2,t} = -\frac{1}{16}\left(Y_{2,t-1} - 8Y_{1,t-1}\right) + \epsilon_{2,t} \tag{1067}$$

In matrix notation

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{1068}$$

or more compactly

$$\Delta Y_t = \Pi Y_{t-1} + \epsilon_t \tag{1069}$$

where $\Delta Y_t = \begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix}$; $\Pi = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix}$ and $\epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$

Separating out the adjustment coefficient and the equilibrium coeffcient as $\alpha' = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$ $\beta' = \begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix}$; equilibrium combination $\beta' Y_t = \xi_t$ and $\Pi = \alpha\beta'$

Putting all these elements the above equations can be written as

$$\Delta Y_t = \Pi Y_{t-1} + \epsilon_t \quad or \quad \Delta Y_t = \alpha\beta' Y_{t-1} + \epsilon_t \tag{1070}$$

Here the VAR system

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{1071}$$

Can be written slightly differently in error correction form as

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{1072}$$

LHS ~$I(0)$ α β' coin vector Level ~$I(1)$ noise ~$I(0)$

The LHS is stationary, noise is stationary, $Y_{1,t}$ and $Y_{2,t}$ as $I(1)$; therefore to balance the time series properties of LHS and RHS β' must be the cointegrating vector that give the linear combination of $I(1)$ variables $Y_{1,t}$ and $Y_{2,t}$ that is stationary.

$$\beta' Y_t = \begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} = Y_{1,t-1} - \frac{1}{8}Y_{2,t-1} \tag{1073}$$

Now the write the ECM in VAR form:

$$Y_{1,t} = \frac{1}{2}Y_{1,t-1} + \frac{1}{16}Y_{2,t-1} + \epsilon_{1,t} \tag{1074}$$

$$Y_{2,t} = \frac{1}{2}Y_{1,t-1} + \frac{15}{16}Y_{2,t-1} + \epsilon_{2,t} \tag{1075}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & \frac{15}{16} \end{pmatrix}\begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} \tag{1076}$$

The relation between ECM and VAR coefficient is:

$$\Pi = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix} = \Pi_1 - I = \begin{pmatrix} \frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & \frac{15}{16} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1077}$$

Number of non-zero eigen values is the rank.
$\begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} Y_{1,t-1} = Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}$
Iterate one period forward and substitute for VAR
$\xi_t = Y_{1,t} - \frac{1}{8}Y_{2,t} = \frac{1}{2}Y_{1,t-1} + \frac{1}{16}Y_{2,t-1} + \epsilon_{1,t} - \frac{1}{8}\left(\frac{1}{2}Y_{1,t-1} + \frac{15}{16}Y_{2,t-1} + \epsilon_{2,t}\right)$
$= \frac{1}{2}Y_{1,t-1} - \frac{1}{16}Y_{1,t-1} + \frac{1}{16}Y_{2,t-1} - \frac{15}{16\times 8}Y_{2,t-1} + \epsilon_{1,t} - \frac{1}{8}\epsilon_{2,t}$
$= \frac{7}{16}Y_{1,t-1} - \frac{7}{16\times 8}Y_{2,t-1} + \epsilon_{1,t} - \frac{1}{8}\epsilon_{2,t} = \frac{7}{16}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \epsilon_{1,t} - \frac{1}{8}\epsilon_{2,t}$
$\xi_t = \frac{7}{16}\xi_{t-1} + \epsilon^*$
where $\epsilon^* = \epsilon_{1,t} - \frac{1}{8}\epsilon_{2,t}$. as $\left|\frac{7}{16}\right| < 1$ ξ_t is a stationary process. Any other scaling does not make any difference.

In $\Pi = \alpha\beta'$if β is scaled by k the α must be scaled by k^{-1} $\Pi = \alpha k^{-1}\beta' k = \alpha\beta$.

Error correction coefficient cannot be uniquely determined, a normalisation can be chosen to help in the economic context.
$\begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\left(\frac{1}{100}\right)100\begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix}$

Notice in $\Pi = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix}$ the row2 equals (-1) times the first row and the second column equals the (-1) the first column. Π does not have independent rows and columns, it is less than full rank. It has rank 1 through matrix is 2×2.Thus these vectors have linear combination resulting in zero. $|\Pi| = -\frac{1}{2}\left(-\frac{1}{16}\right) - \frac{1}{2}\left(\frac{1}{16}\right) = 0$. One of the eigen values is zero.
$|\Pi - vI| = \begin{vmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{vmatrix} - \begin{vmatrix} v & 0 \\ 0 & v \end{vmatrix} = \begin{vmatrix} -\frac{1}{2} - v & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} - v \end{vmatrix}$
$\begin{vmatrix} -\frac{1}{2} - v & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} - v \end{vmatrix} = \left(-\frac{1}{2} - v\right)\left(-\frac{1}{16} - v\right) - \frac{1}{2}\left(\frac{1}{16}\right) = \frac{1}{32} + \frac{9}{16}v + v^2 - \frac{1}{32} = \frac{9}{16}v + v^2 = 0$
$v\left(\frac{9}{16} + v\right) = 0$; Thus two roots of the system are $v_1 = 0$; $v_2 = -\frac{9}{16}$
Problems in Engle-Granger Methodology

1. There are two major problems with the Engle-Granger methodology of testing cointegration.
2. First, ordering of variables may influence inference about the stationarity of the residuals in the small sample though this may be eliminated in the large samples. Secondly, it relies on two step procedure.
3. Step 1 involves getting the residual from the original regression.

4. Step 2 is to test the stationarity of the residual. $\Delta e_t = a_1 e_{t-1} + v_t$ Here a_1 is estimated using the residual in the previous estimation.

5. Thus any problem occurring in that estimation also carry on to the next step.

Estimationof Multiple Cointegrating Vectors Example 2 Johansen procedure allows simultaneous estimation of multiple cointegration vectors and also allows to estimate the restricted version of cointegration and speed of adjustment.

ECM	Equilibrium
$\Delta Y_{1,t} = -\frac{1}{2}\xi_{1,t-1} + \frac{1}{4}\xi_{2,t-1} + \epsilon_{1,t}$	$\xi_{1,t} = Y_{1,t} - \frac{1}{8}Y_{2,t}$
$\Delta Y_{2,t} = \frac{1}{8}\xi_{1,t-1} - \frac{5}{8}\xi_{2,t-1} + \epsilon_{1,t}$	$\xi_{2,t} = Y_{1,t} - \frac{1}{4}Y_{2,t}$
$\Delta Y_{3,t} = \frac{1}{4}\xi_{1,t-1} + \frac{3}{8}\xi_{2,t-1} + \epsilon_{3,t}$	

Estimation of Multiple Cointegrating Vectors: Example 2
With three variables it is possible to define two equilibrium relations.

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \\ \Delta Y_{3,t} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & -\frac{5}{8} \\ \frac{1}{4} & \frac{3}{8} \end{pmatrix} \begin{pmatrix} \xi_{1,t-1} \\ \xi_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix} \tag{1078}$$

Inserting the long run matrix it becomes:

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \\ \Delta Y_{3,t} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{8} & -\frac{5}{8} \\ \frac{1}{4} & \frac{3}{8} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{8} & 0 \\ 0 & 1 & -\frac{1}{4} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix} \tag{1079}$$

$$\alpha = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \\ \frac{1}{4} & \frac{3}{8} \end{pmatrix}; \beta = \begin{pmatrix} 1 & -\frac{1}{8} & 0 \\ 0 & 1 & -\frac{1}{4} \end{pmatrix};$$

$$\Pi = \alpha\beta' = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \\ \frac{1}{4} & \frac{3}{8} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{8} & 0 \\ 0 & 1 & -\frac{1}{4} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{5}{16} & -\frac{1}{16} \\ \frac{1}{8} & -\frac{41}{64} & \frac{5}{32} \\ \frac{1}{4} & -\frac{11}{32} & -\frac{3}{32} \end{pmatrix} \tag{1080}$$

Eigen value of Π: $\begin{pmatrix} -0.18 & -0.62 & -0.42 \end{pmatrix}$
There are thre cointegrating vectors.

Cointegration and Error Correction: Example 3 Example 3: More lags
ECM Form

$$\Delta Y_{1,t} = -\frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \frac{1}{8}\Delta Y_{1,t-1} + \frac{1}{4}\Delta Y_{2,t-1} + \epsilon_{1,t} \tag{1081}$$

$$\Delta Y_{2,t} = \frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \frac{1}{4}\Delta Y_{1,t-1} - \frac{3}{4}\Delta Y_{2,t-1} + \epsilon_{2,t}$$

In matrix notation

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \epsilon_t \tag{1082}$$

$$\Pi = \alpha\beta' = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{8} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix}; \Gamma_1 = \begin{pmatrix} \frac{1}{8} & \frac{1}{4} \\ \frac{1}{4} & -\frac{3}{4} \end{pmatrix} \tag{1083}$$

$Y_{1,t}$ and $Y_{2,t}$ must be cointegrated to maintain stationarity in LHS and RHS.
Write this in VAR form

$$Y_{1,t} = \frac{5}{8} Y_{1,t-1} + \frac{5}{16} Y_{2,t-1} - \frac{1}{8} Y_{1,t-2} - \frac{1}{4} Y_{2,t-2} + \epsilon_{1,t} \tag{1084}$$

$$Y_{2,t} = \frac{3}{4} Y_{1,t-1} + \frac{13}{16} Y_{2,t-1} - \frac{1}{4} Y_{1,t-2} + \frac{3}{4} Y_{2,t-2} + \epsilon_{2,t} \tag{1085}$$

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \epsilon_t \tag{1086}$$

$\Pi_1 = \begin{pmatrix} \frac{5}{8} & \frac{5}{16} \\ \frac{3}{4} & -\frac{3}{16} \end{pmatrix}; \Pi_2 = \begin{pmatrix} -\frac{1}{8} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{pmatrix}$. the coefficient of VAR and ECM are related as before as:

$\Pi = \Pi_1 + \Pi_2 - I \quad ; \quad \Gamma_1 = -\Pi_2;$

$$\begin{pmatrix} -\frac{1}{2} & \frac{1}{16} \\ \frac{1}{2} & -\frac{1}{16} \end{pmatrix} = \begin{pmatrix} \frac{5}{8} & \frac{5}{16} \\ \frac{3}{4} & \frac{3}{16} \end{pmatrix} + \begin{pmatrix} -\frac{1}{8} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{3}{4} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{1087}$$

eigne $=[-0.5625, 0]$
Generalisation of a VAR model

$$Y_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + ... + \Pi_p Y_{t-p} + \epsilon_t \tag{1088}$$

ECM form

$$\Delta Y_t = \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + +\Gamma_2 \Delta Y_{t-2} + ... + +\Gamma_p \Delta Y_{t-p-1} + \epsilon_t \tag{1089}$$

$\Pi = \Pi_1 + \Pi_2 + ... + \Pi_p - I \quad ; \quad \Gamma_i = -\left(\Pi_{i+1} + \Pi_{i+2} + ... + \Pi_p\right)$ for $i = 1, ..., p-1$.

5.0.7 Determinants, characteristic roots and Trace

Let A be $N \times N$ matrix. Then the determinants of $|A| = \prod_{i=1}^{N} \lambda_i$ for i $=\lambda_1, \lambda_2, ..., \lambda_n$
$|A - \lambda I| = 0$ implies
$(a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) .... (a_{nn} - \lambda) = 0$
$\lambda^n + b_1 \lambda^{n-1} + b_2 \lambda^{n-2} + ... + b_{n-1}\lambda + b_n = 0$
by factor rule of the polynomial $\prod_{i=1}^{N} \lambda_i = (-1)^n b_n = |A| = \lambda_1 \lambda_2 \lambda_3 ..... \lambda_{n-1} \lambda_n$

Rank of A equals the number of non-zero characteristic roots. If $|A| \neq 0$ then non of $\lambda_i = 0$. A has full rank. If rank of A is zero then each element of A must be zero and $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_{n-1} = \lambda_n = 0$. In intermediate case rank of A is between 0 and N.

Illustrations:

$|A| = \begin{vmatrix} 0.5 & -1 \\ -0.2 & 0.4 \end{vmatrix}$ $\lambda_1 = 0.9$ and $\lambda_2 = 0$ rank of A is 1.

$|A| = \begin{vmatrix} 0.5 & -0.2 \\ -0.2 & 0.5 \end{vmatrix}$ $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$ rank of A is 2. Eigen vectors $V_1 = \begin{pmatrix} -0.7071 \\ -0.7071 \end{pmatrix}$; $V_2 = \begin{pmatrix} -0.7071 \\ 0.7071 \end{pmatrix}$

Check eigen vectors are orthogonal, $(V_1)^{'}(V_2) = \begin{pmatrix} -0.7071 & -0.7071 \end{pmatrix} \begin{pmatrix} -0.7071 \\ 0.7071 \end{pmatrix} = 0$

$(V_1V_2)^{'}(V_1V_2) = \begin{pmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{pmatrix}^{'} \begin{pmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{pmatrix} = (V_1V_2)(V_1V_2)^{'} = I =^{'} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Determinants, characteristic roots and Trace

$|A| = \begin{vmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.5 \end{vmatrix}$ $\lambda_1 = 0.9$, $\lambda_2 = 0.3$, $\lambda_3 = 0.3$ and rank of A is 3.

$|A| = \begin{vmatrix} 0.5 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0.4 \\ -0.25 & -0.1 & -0.1 \end{vmatrix}$ $\lambda_1 = 0.4$, $\lambda_2 = 0.4$, $\lambda_3 = 0$ Rows are linearly dependent and rank of A is 1.

Stability of VAR

$X_t = A_0 + A_1X_{t-1} + \varepsilon_t$

for stability check the homegenous part of the solution

$X_t = A_1X_{t-1}$

use the undetermined coefficient to solve this problem.

$X_{i,t} = C_i\lambda^t$

Insert these in the extended VAR

$C_1\lambda^t = a_{11}C_1\lambda^{t-1} + a_{12}C_2\lambda^{t-1} + a_{13}C_3\lambda^{t-1} + ... + a_{1,n}C_n\lambda^{t-1}$

$C_2\lambda^t = a_{21}C_1\lambda^{t-1} + a_{22}C_2\lambda^{t-1} + a_{23}C_3\lambda^{t-1} + ... + a_{2,n}C_n\lambda^{t-1}$

..

$C_n\lambda^t = a_{n1}C_1\lambda^{t-1} + a_{n,2}C_2\lambda^{t-1} + a_{n,3}C_3\lambda^{t-1} + ... + a_{n,n}C_n\lambda^{t-1}$

For homogenous case

$C_1(a_{11} - \lambda) + a_{12}C_2 + a_{13}C_3 + ... + a_{1,n}C_n = 0$

$a_{21}C_1 + C_2(a_{22} - \lambda) + a_{23}C_3 + ... + a_{2,n}C_n = 0$

..

$a_{n1}C_1 + a_{n,2}C_2 + a_{n,3}C_3 + ... + C_n(a_{nn} - \lambda) = 0$

Stability of VAR

$$\begin{bmatrix} (a_{11} - \lambda) & a_{12} & a_{13} & a_{1n} \\ a_{21} & (a_{11} - \lambda) & . & a_{2,n} \\ . & . & . & . \\ a_{n1} & a_{n,2} & . & (a_{nn} - \lambda) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ . \\ C_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Non trivial solution requires

$$\begin{vmatrix} (a_{11}-\lambda) & a_{12} & a_{13} & a_{nn} \\ a_{21} & (a_{11}-\lambda) & . & a_{2,n} \\ . & . & . & . \\ a_{n1} & a_{n,2} & . & (a_{nn}-\lambda) \end{vmatrix} = 0$$

The determinant will be nth order polynomial and n values of $\lambda_1, \lambda_2, ..., \lambda_n$. Necessary and sufficient condition for stability is that all characteristic roots lie within the unit circle.

Stability of VAR and Johansen Procedure This now should help to understand the Johansen Procedure in cointegration. As mentioned before.

Start with X_t be vector of $N \times 1$ dimension each integrated of order 1 $I(1)$. The VAR is given by

$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + ... + \Pi_k X_{t-k} + \varepsilon_t$

Π_i is $N \times N$ matrix of coefficients This implies

$$\begin{pmatrix} X_{1,t} \\ . \\ . \\ X_{n,t} \end{pmatrix} = \begin{pmatrix} \Pi_{111} & \Pi_{112} & . & . & . & \Pi_{11N} \\ \Pi_{121} & . & . & . & . & \Pi_{121} \\ . & . & . & . & . & . \\ \Pi_{1N1} & \Pi_{N21} & . & . & . & \Pi_{1NN} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ . \\ . \\ X_{n,t-1} \end{pmatrix} + ... + \begin{pmatrix} \Pi_{11k} & \Pi_{12k} & . & . & . & \Pi_{1Nk} \\ \Pi_{21k} & . & . & . & . & \Pi_{21k} \\ . & . & . & . & . & . \\ \Pi_{N1k} & \Pi_{N2k} & . & . & . & \Pi_{NNk} \end{pmatrix} \begin{pmatrix} X \\ . \\ . \\ X \end{pmatrix}$$

ECM representation of this VAR is given by

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1090}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_1 + ... + \Pi_i$ for $i = 1..k$

Γ_k gives the long run solution.

Here the terms in LHS and the first $k-1$ elements of difference of X_t are $I(0)$. The last element $\Gamma_k X_{t-k} = \alpha\beta' X_{t-k}$ is the linear combination of $I(1)$ variables. it must be stationary to balance the time series properties of right and left hand side. It has $N-r$ ranks.

Johansen uses the canonical correlation method to produce all the distinct combinations of level of X which produce high correlation with $I(0)$ elements in above equations.

These correlates are cointegrating vectors. Johansen develops maximum likelihood method to estimate distinct cointegrating vectors, devises how the maximum likelihood ratio test could be applied to decide the significance of these cointegrating vectors.

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \alpha\beta' X_{t-k} + \varepsilon_t \tag{1091}$$

Likelihood function

$$L(\alpha, \beta, \Delta, \Gamma_1..\Gamma_K) = (\Omega)^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \left(e_t \Omega^{-1} e_t' \right) \right\} \tag{1092}$$

Where T is the number of observations and Ω the covariance matrix of e_t. The ECM can be written as:

$$\Delta X_t + \alpha\beta' X_{t-k} = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \varepsilon_t \tag{1093}$$

Johansen first proposes correcting the effects of k lags of ΔX_t on ΔX_t and it level X_{t-k} . This is done by first regressing ΔX_t on lagged differences of ΔX_t on the RHS and getting the residuals R_{0t} . Next regress X_{t-k} on on lagged differences of ΔX_t on the RHS and getting the residuals R_{kt} .

These two sets of residuals are related as $e_t = R_{0t} + \alpha\beta' R_{kt}$

Now write the maximum likelihood in terms of these residuals

$$L(\alpha, \beta, \Omega) = (\Omega)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\left[\left(R_{0t} + \alpha\beta' R_{kt}\right)' \Omega^{-1}\left(R_{0t} + \alpha\beta' R_{kt}\right)\right]\right\} \tag{1094}$$

If β is known (from the long run equilibrium relations) estimate α and Ω by regressing R_{0t} on $\beta' R_{kt}$. Thus $\widehat{\alpha}$ and $\widehat{\Omega}$ can be expressed as

$$R_{0t} = -\alpha\beta' R_{kt} + \varepsilon_t \tag{1095}$$

Define the sum of the product of residuals as: $S_{i,j} = T^{-1}\sum_{t=1}^{T} R_{i,t} R'_{i,t}$ for $i, j = o, k$

Now the estimators of $\widehat{\alpha}$ and $\widehat{\Omega}$ are given by

$$\widehat{\alpha}\ (\beta) = \left(\beta R'_{k,t} R_{kt} \beta'\right)^{-1} \left(\beta R_{kt} R_{ot}\right) = -S_{ok}\left(\beta' S_{kk} \beta'\right)^{-1} \tag{1096}$$

$$\widehat{\Omega}\ (\beta) = S_{oo} - S_{ok}\beta\left(\beta' S_{kk}\beta\right)^{-1}\beta' S_{ko} \tag{1097}$$

These estimated parameters can now be substituted to evaluate the likelihood as:

$$L(\beta) = \left|\Omega\left(\beta\right)\right|^{-\frac{T}{2}} = \left|S_{oo} - S_{ok}\beta\left(\beta' S_{kk}\beta\right)^{-1}\beta' S_{ko}\right|^{-\frac{T}{2}} \tag{1098}$$

Finding the cointegration vector involves finding vector β that minimises this function.

$$F = \left|S_{oo} - S_{ok}\beta\left(\beta' S_{kk}\beta\right)^{-1}\beta' S_{ko}\right|^{-\frac{T}{2}} \tag{1099}$$

Johansen shows how this can be done by solving the eigenvalues and finding the eigen vectors using canonical correlations. $\widehat{\beta}$ is a set of eigenvector estimated together with $N-1$ eigenvalues $\widehat{\lambda}$. Column of $\widehat{\beta}$ is significant only if the corresponding eigenvalue is significantly different from zero. Let $\widehat{\lambda}_i$ be ordered by its value as $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \widehat{\lambda}_3 > .... > \widehat{\lambda}_{N-1}$. Similarly let $\widehat{\beta}$ be ordered by the corresponding eigen values

The maximum likelihood

$$\widehat{\Omega}\ (\beta) = \left|S_{oo}\right| \prod_{i=1}^{N}\left(1 - \widehat{\lambda}_i\right) \tag{1100}$$

Testing for r cointegrating vectors H_o: $\lambda_i = 0, i = r+1, ..., N$.is equivalent to testing that first r eigenvalues are non-zero.

Then there is restricted estimation of Δ given by

$$\widehat{\Omega}\ (\beta) = \left|S_{oo}\right| \prod_{i=1}^{r}\left(1 - \widehat{\lambda}_i\right) \tag{1101}$$

The likelihood ratio test is made from these unrestricted and restricted estimates as defined below:

$$LR(N-r) = -2\ln Q = -T\sum_{t=1}^{T}\ln\left(1 - \widehat{\lambda}_i\right) \tag{1102}$$

where
$$Q = \frac{\text{Restricted maximised likelihood}}{\text{Unrestricted maximised likelihood}} \tag{1103}$$

$LR(N-r)$ has degrees of freedom equal to the number of restrictions

for $\widehat{\lambda}_i = 0$ $i = r+1, ..., N; LR(N-r) = 0$. Here $LR(N-r)$ tend to get large as one or more of $\widehat{\lambda}_i$ approach to unity. LR does not have χ^2 distribution

Brownian motion theory is used to find its asymptotic distributions. Johansen has tabulated these values for VAR without constant, VAR with restricted constant (only the cointegrating vectors) and VAR with unrestricted constants.

Theorem: The maximum likelihood estimates of the space spanned by β is the space spanned by the r canonical variates corresponding to the r largest squared canonical correlations between the residual of ΔX_t and it level X_{t-k} corrected for the effect of the lagged differences of the X process.

The likelihood ratio test for that there are at most r cointegrating vectors $LR(N-r) = -2\ln Q = -T\sum_{t=1}^{T} \ln\left(1-\widehat{\lambda}_i\right)$ .The asymptotic distribution of the maximum likelihood ratio is the function of $(N-r)$ dimensional Brownian motion.

A set of critical values are tabulated which are correct for all models, space spanned by β and $\widehat{\beta}$ are the same. The maximum likelihood estimates of β is obtained as eigen vectors corresponding to the largest eigen values by solving equation

$\left|\lambda S_{kk} - S_{ko}S_{oo}^{-1}S_{ok}\right| = 0$

This given n eigen values $\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3, ..., \widehat{\lambda}_{N-1}$ and corresponding eigen vectors $\widehat{v}_1, \widehat{v}_2, \widehat{v}_3, ..., \widehat{v}_{N-1}$ and the $\widehat{\beta} = (\widehat{v}_1, \widehat{v}_2, \widehat{v}_3, ..., \widehat{v}_{N-1})$

the eigen values are the largest squared canonical correlations between level of residuals R_{kt} and R_{ot}.

5.0.8 Cannonical correlation: an example

Consider a 4×4 correlation matrix of the following form $R = \begin{bmatrix} 1 & 0.4 & 0.5 & 0.6 \\ 0.4 & 1 & 0.3 & 0.4 \\ 0.5 & 0.3 & 1 & 0.2 \\ 0.6 & 0.4 & 0.2 & 1 \end{bmatrix}$. This is a symmetric matrix. Split the matrix in four equal parts. $\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$ where $R_{11} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$ $R_{22} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$; $R_{12} = \begin{pmatrix} 0.5 & 0.6 \\ 0.3 & 0.4 \end{pmatrix}$ and $R_{21} = \begin{pmatrix} 0.5 & 0.3 \\ 0.6 & 0.5 \end{pmatrix}$

Set the determinant of the correlation matrix R to find the roots that correspond to highest correlation among its components

$$|R| = \begin{vmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{vmatrix} = R_{11}R_{22} - R_{12}R_{21} = 0 \tag{1104}$$

This implies
$$I = R_{22}^{-1}R_{12}R_{11}^{-1}R_{21} \tag{1105}$$

$$
\begin{aligned}
RR &= R_{22}^{-1}R_{12}R_{11}^{-1}R_{21} = \\
&\quad \begin{pmatrix} 1.041 & -0.208 \\ 0.208 & 1.041 \end{pmatrix}\begin{pmatrix} 0.5 & 0.6 \\ 0.3 & 0.4 \end{pmatrix} \\
&\quad \begin{pmatrix} 1.190 & -0.476 \\ -0.476 & 1.190 \end{pmatrix}\begin{pmatrix} 0.5 & 0.3 \\ 0.6 & 0.4 \end{pmatrix} \\
&= \begin{pmatrix} 0.2063 & 0.2501 \\ 0.2778 & 0.3403 \end{pmatrix} \qquad (1106)
\end{aligned}
$$

Find the eigen values

$$
|RR - vI| = \begin{vmatrix} 0.2063 - v & 0.2501 \\ 0.2778 & 0.3403 - v \end{vmatrix} = 0 \qquad (1107)
$$

This implies
$(0.2063 - v)(0.3403 - v) - (0.2778)(0.2501) = 0.0702 - .2063v - .3403v + v^2 - 0.0695$

$$
v^2 - 0.547v + 0.007 = 0 \qquad (1108)
$$

$$
v_1, v_2 = \frac{0.547 \pm \sqrt{(0.547)^2 - 4(0.007)}}{2} = 0.534; 0.013 \qquad (1109)
$$

It solves for $v_1 = 0.534$ and $v_2 = 0.013$

Canonical correlation is the square root of eigenvalues $R_{c1} = \sqrt{v_1 = 0.534} = 0.731$ and $R_{c1} = \sqrt{v_2 = 0.013} = 0.11$

Johansen applied this idea on correlation matrices R_{0t} and R_{kt}. The highest eigen value are associated to the zero value of canonical product matrix as illustrated above. Johansen test is a reduced rank test for non-stationarity part

$\widehat{\lambda}_i = 0 \; i = r+1, ..., N$;

$LR(N - r) = 0$.

In other word testing $r = 1$ is testing for $\widehat{\lambda}_2 = \widehat{\lambda}_3 = ... = \widehat{\lambda}_{N-1} = 0$ whereas $\widehat{\lambda}_1 > 1$

Summary on Steps

1. estimate the VAR in first differences get the errors for each equation

$$
\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1}\Delta X_{t-k+1} + e_{1,t} \qquad (1110)
$$

2. Then regressed the lagged term in lagged differences and get the errors for each equation

$$
X_{t-1} = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1}\Delta X_{t-k+1} + e_{2,t} \qquad (1111)
$$

3. Compute the canonical correlations between $e_{1,t}$ and $e_{2,t}$

$$
\left| \lambda_i S_{22} - S_{12} S_{11}^{-1} S_{12}^{'} \right| = 0 \qquad (1112)
$$

$S_{ii} = T^{-1}\sum_{t=1}^{T} e_{it}e_{it}^{'}$, $S_{12} = T^{-1}\sum_{t=1}^{T} e_{2t}e_{2t}^{'}$

4. MLE of cointegrating vectors are n columns that are non-trivial solutions for

$$\lambda_i S_{22}\pi_i = S_{12}S_{11}^{-1}S_{12}^{'}\pi_i \tag{1113}$$

See Enders(425).

References

[1] Apergis N, O. C. Dincer and J. E. Payne (2010) The relationship between corruption and income inequality in U.S. states: evidence from a panel cointegration and error correction model, Public Choice, 145, 1/2,125-135

[2] Banerjee A, J Dolado and J W Galbraith and D. F. Hendry (1993) Cointegration, error correction and the econometric analysis of Non-stationary Data, Oxford.

[3] Bansal R and D Kiku (2011) Cointegration and Long-Run Asset Allocation, Journal of Business & Economic Statistics, 29, 1, 161-173

[4] Balke N. S and T. B. Fomby (1997) Threshold Cointegration, International Economic Review, 38, 3, 627-645

[5] Bhattarai, K. (2008) An Empirical Study of Interest Determination Rules, *Applied Financial Economics*, 18:4:327-343.

[6] Campbell J. Y. and R. J. Shiller (1987) Cointegration and Tests of Present Value Models Journal of Political Economy, 95, 5, 1062-1088

[7] Davidson J and S. Hall (1991) Cointegration in Recursive Systems, Economic Journal, 101, 405 , 239-251

[8] Dickey D.A. and W.A. Fuller (1979) Distribution of the Estimator for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74:366:427-437,June.

[9] Hendry D.F and J. A. Doornik (1994) Modelling linear dynamic econometric Systems, Scottish Journal of Polifical Economy, 41. I, 1-33

[10] Doornik J A and D.F. Hendry (2012) PC-Give Volumes I-III, GiveWin Timberlake Consultants

[11] Enders W (1988) Arima and Cointegration Tests of PPP under Fixed and Flexible Exchange Rate Regimes, Review of Economics and Statistics, 70, 3, pp. 504-508

[12] Enders W and P. L.Siklos (2001) Cointegration and Threshold Adjustment, Journal of Business & Economic Statistics, 19, 2,166-176

[13] Enders W. (2010) Applied Econometric Time Series, 3rd ed., Willey.

[14] Engle R E and C.W.J. Granger (1987) Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica, 55, 2, 251-276.

[15] Greene W. (2008) Econometric Analysis, Sixth Edition, Pearson Education, New York.

[16] Hall A. D and H. M. Anderson, C. W. J. Granger (1992) A Cointegration Analysis of Treasury Bill Yields, Review of Economics and Statistics, 74, 1,116-126

[17] Hansen, B. E. (1992). Tests for Parameter Instability in Regressions with I(1) Processes, Journal of Business and Economic Statistics, 10, 321-335.

[18] Hamilton James D. (1994) Time Series Analysis, Princeton.

[19] Harris R and R Sollis (2003) Applied Time Series Modelling and Forecasting, John Wiley and Sons.

[20] Harris R(1995) Using cointegration analysis in econometric modelling, Prentice Hall.

[21] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[22] Hualde J (2006) Unbalanced Cointegration, Econometric Theory, 22, 5, 765-814

[23] Hylleberg S and G. E. Mizon (1989) Cointegration and Error Correction Mechanisms, Economic Journal, 99, 395,113-125

[24] Jawadi F, C. Bruneau and N. Sghaier (2009) Nonlinear Cointegration Relationships between Non-Life Insurance Premiums and Financial Markets, Journal of Risk and Insurance, 76, 3, 753-783

[25] Johansen Soren (1988) Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control, 12:231-254, North Holland.

[26] Kim Y(1990) Purchasing Power Parity in the Long Run: A Cointegration Approach Journal of Money, Credit and Banking, 22, 4 , 491-503

[27] Kuthbertson K, S. G. Hall and MP Taylor (1992) Applied Econometric Techniques, University of Michigan Press.

[28] Nelson C. R. and C. I. Plosser (1982) Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications, Journal of Monetary Economics, 10, 139-163.

[29] Park, J. Y. (1992). Canonical Cointegrating Regressions, Econometrica, 60, 119-143.

[30] Phillips P. C. B. and S. Ouliaris (1990) Asymptotic Properties of Residual Based Tests for Cointegration, Econometrica, 58, 1,165-193

[31] Phillips P.C.B. (1987) Time Series Regression with an Unit Root, Econometrica, 55, 2, 277-301.

[32] Qu Z and P. Perron (2007) A Modified Information Criterion for Cointegration Tests Based on a VAR Approximation, Econometric Theory, 23, 4, 638-685

[33] von Hagen J (1989) Relative Commodity Prices and Cointegration, Journal of Business & Economic Statistics, 7, 4, 497-503

[34] Solo V (1995) Errors in Variables and Cointegration, Econometric Theory, 11, 1, 60-80

[35] Trapani L and G Urga (2010) Micro versus macro cointegration in heterogeneous panels, Journal of Econometrics 155 1 18

[36] Turner P (2009) Testing for Cointegration Using the Johansen Approach: Are We Using the Correct Critical Values?, Journal of Applied Econometrics, 24, 5, 825-831

[37] Villani M (2005) Bayesian Reference Analysis of Cointegration, Econometric Theory, 21, 2, 326-357

[38] Westerlund J (2008) Panel Cointegration Tests of the Fisher Effect Panel, Journal of Applied Econometrics, 23, 2,193-233

6 L5: Vector Autoregression

When it is not possible to isolate exogenous and endogenous variables, it is natural to assume that path of y_t affected by another variable z_t and the path of z_t affected by another variable y_t . Their path is determined simultaneously. Unrestricted VAR and restricted VAR have been widely used analysing time series properties of a system of variables. An indicative literature on VAR modeling:

- Sims (1980), Lucas and Sargent (1981), Bernanke (1986), Judge, Hill, Griffiths, Lutkepohl and Lee (1988), Pagan and Wickens (1989), Wallis (1989), Hendry and Doornik (1994), Kocherlakota and Yi (1996), Kadiyala and Karlsson (1997), Sims and Zha (1999), Patterson (2000), Mills (2000), Stock, and Watson (2001), Canova and Nicoló (2002),Garratt, Lee, Pesaran, and Shin (2003), Harris and Sollis (2003), Davidson and MacKinnon (2004), Mountford (2005), Kapetanios, George, Pagan, and Scott (2007), Athanasopoulos and Vahid (2008), Vargas-Silva (2008), Benati and Surico (2009), Mertens and Ravn (2010), Fry and Pagan (2011), Bhattarai and Mallick (2013)

- Hendry (1974), Hendry (2002), Stock and Watson (2005), Swensen (2006), Enders and Hurn (2006), Dowd and Blake (2006), Dees, Mauro, Pesaran and Smith (2007), Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), Fanelli (2007), Rafiq and Mallick (2008), Dungey and Fry (2009), Mertens and Ravn (2010), Chahrour, Schmitt-Grohé, and Uribe (2012), Favero, and Giavazzi (2012), Phillips and Magdalinos (2013)

VAR analysis is popular in the econometrics literature (see Sims (1980) Fair (1984) Bernanke (1986), Wallis (1989), Blanchard and Quah (1989) Hamilton (1994) Hendrry(1995) Kocherlakota and Yi (1996) Patterson (2000), Davidson and MacKinnon (2004) Enders (2010)).

Bhattarai and Mallick (2013) apply a VAR opn time series data of the China and the US on wages, interest rates, exchange rates, GDP, current account balance and the US trade decifit to find empirical evidence on above analysis. A structural VAR model estimated in line of Sim (1980) and Bernanke(1986) with restrictions appropriate to theoretical derivations (see Fry and Pagan (2011) for up to date review on this). We limit our analysis to five variables that include relative wage between China and the US (w_{cu}), interest rate differential between China and the US (r_{cu}), Chinese real effective exchange rate (e), GDP of China relative to that of the US (ry_{cu}) and the current account balance (CA_u) determining the ordering of these variables in the SVAR following logics explained in Rafiq and Mallick (2008). In a nutshell we try to show how the relative prices of labour, capital and the currency affect the economic activities in China and trade balance in the US. The raw time series of these data are presented in Figure 1. When the Chinese economy has been growing rapidly, the exchange rate being fixed leads us to use China's real exchange rate ($e_{c,t}$) rather than the nominal exchange rate. By doing so, we are also capturing the relative price effects. China's unit labour cost (ULC) is measured as total wage bill over real output (nominal output divided by CPI (1985=100)). Then relative wage ($w_{cu,t}$) is calculated by dividing ULC-China over ULC-US. Relative GDP ($ry_{cu,t}$) on the other hand has been defined as Chinese GDP in dollar terms over US GDP. We calculate interest rate differential ($r_{cu,t}$) as the difference between Chinese average inter-bank rate and US 3-month Tbill rate. Current account balance ($CA_{u,t}$) for the US is used as the percentage of US nominal GDP. With these five variables, we formulate a first-order structural VAR of the following form:

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t-1} \\ r_{cu,t-1} \\ e_{c,t-1} \\ ry_{cu,t-1} \\ CA_{u,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{ryt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1114}$$

where matrix notations can be employed for more compact representation.

$$X_t = \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix}; \; X_{t-1} = \begin{bmatrix} w_{cu,t-1} \\ r_{cu,t-1} \\ e_{c,t-1} \\ ry_{cu,t-1} \\ CA_{u,t-1} \end{bmatrix}; \; \varepsilon_t = \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{ryt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1115}$$

Thus the path of X_{it} is affected by both contemporaneous and lagged effects of X_{jt} as measured by Γ_0 and Γ_1 and its own past values.Consider

$$X_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 X_{t-1} + B^{-1}\varepsilon_t \tag{1116}$$

$$B^{-1} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix}; \; \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix}$$

The reduced form of this VAR system is then given by:

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1117}$$

where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$, $e_t = B^{-1}\varepsilon_t$

Reduced form is estimated with the available data; then structural shocks are retrieved using $e_t = B^{-1}\varepsilon_t$.This requires estimation of the variance covariance matrix of the error term:

$$\sum = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix} \tag{1118}$$

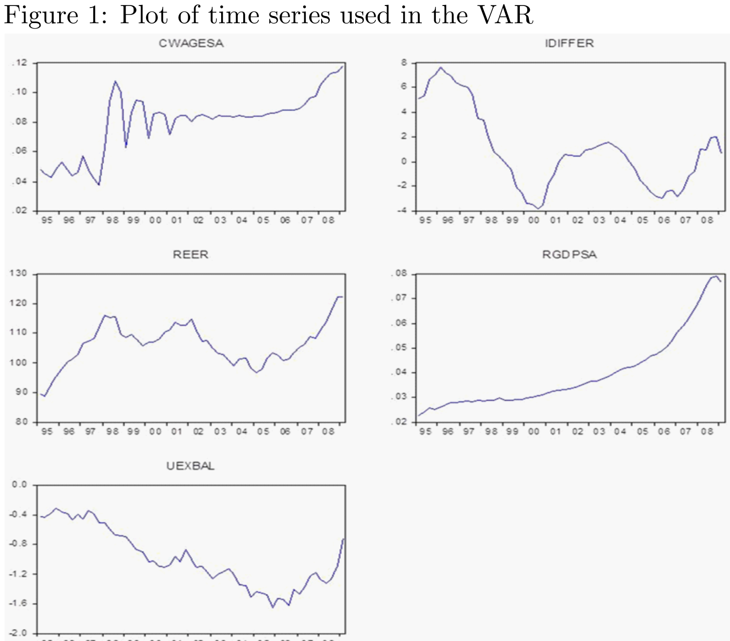
where $\sigma_{ij} = \frac{1}{T}\sum_{t=1}^{T} e_{ij}e_{ij}^{'}$.

VAR is a-theoretic. In order to understand the long-run dynamics, we perform impulse response shock analysis, as the results from impulse responses are more informative than the estimated VAR regression coefficients (see Stock and Watson, 2001). It is customary to impose restrictions on coefficients based on prior economic theory. These restrictions can be on parameters, variance covariance matrices or symmetry.

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1119}$$

Quarterly observations from 1995-Q1 to 2009-Q1 are used to estimate the model with two optimal lags. All the data have been gathered from Datastream and the variables are plotted in Figure 1. Since there is evidence of a structural break around 1994Q1 in China (see for example Baak (2008)), our sample in this paper starts from 1995Q1. Furthermore there is unavailability of quarterly data for the variables involved in this paper prior to 1995Q1.

Figure 1: Plot of time series used in the VAR



6.0.9 Impulse response Analysis

The VAR is formulated with the following ordering: relative wage, interest rate differential, Chinese REER, relative GDP, and US current account balance. Shocks are extracted by applying a recursive identification structure with the above ordering to a vector error correction model.

$$\begin{bmatrix} w_{c,t} \\ r_{c,t} \\ e_{c,t} \\ ry_{u,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \overline{w}_{c,t} \\ \overline{r}_{c,t} \\ \overline{e}_{c,t} \\ \overline{ry}_{u,t} \\ \overline{CA}_{u,t} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \\ e_{3t-i} \\ e_{4t-i} \\ e_{5t-i} \end{bmatrix} \tag{1120}$$

Errors of the reduced form equations are related to the structural parameters as:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \\ e_{5t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1121}$$

Introducing more simplifying assumptions:

$$\begin{bmatrix} w_{c,t} \\ r_{c,t} \\ e_{c,t} \\ pm_{u,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \overline{w}_{c,t} \\ \overline{r}_{c,t} \\ \overline{e}_{c,t} \\ \overline{pm}_{u,t} \\ \overline{CA}_{u,t} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} \end{bmatrix} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1122}$$

where $\phi_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1}$

$\phi_{i,j}(n)$ are impulse response coefficients for equation n. More compactly this can be represented as:

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i(i)\epsilon_{t-i} \tag{1123}$$

Given the higher US imports (more than three times the amount they export), incurring a huge overall trade deficit, there is a growing pressure on China to raise the value of its currency, particularly from the US. This concern can be assessed via a structural VEC exercise whether the deficit is due to relative domestic demand or relative prices (real exchange rate). We therefore have used relative GDP and REER as a relative price variable. The impulse responses of REER shocks on relative GDP show that REER appreciation harms Chinese exports thereby helping US GDP increase faster than Chinese GDP, thereby leading to a decline in relative GDP between the two countries. This suggests that Chinese yuan real appreciation is required to ensure sustainability, as relative GDP shocks only lead to short-run appreciation in REER (see Figure 4). Xu (2008) reports a statistically significant long-run relationship between the RMB/dollar exchange rate and the US trade deficit with China, suggesting a need for China to adjust its exchange rate policy to help reduce the ever mounting US trade deficit.

Table 1: Estimated Matrix of Long-run Impact					
Shocks in $\longrightarrow$	rw	ird	$reer$	ry	cab
rw	0.5387*	-0.3595	-0.7420*	0.9380	-0.1203
ird	0.3064	1.2237*	0.3780	-0.6904	0.1899
$reer$	-0.8826*	0.1650	0.5077	-0.4475	0.2744*
ry	0.0505	0.0432	-0.1436*	0.2028	0.0410
cab	-0.0204	0.0637*	-0.0447	0.0772	0.0466*

:

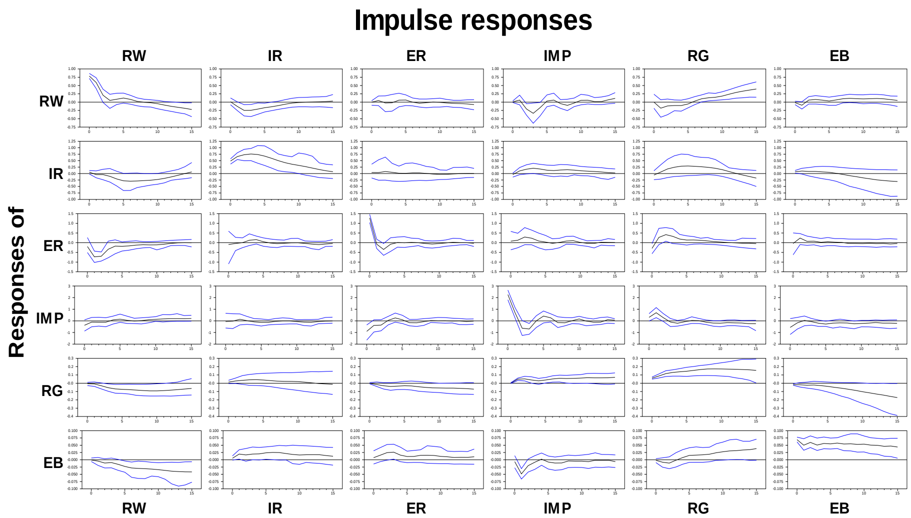
Table 2: Estimated Matrix of Short-run effects					
Shocks in $\longrightarrow$	rw	ird	$reer$	ry	cab
rw	0.8837*	0.0000	0.0000	0.0000	0.0000
ird	0.0547	0.6074*	0.0000	0.0000	0.0000
$reer$	-0.4894*	0.0895	1.5268*	0.0000	0.0000
ry	0.0019	0.0104	-0.0133	0.0766*	0.0000
cab	-0.0074	0.0225*	0.0031	-0.0304*	0.0669*

Table 3: Variance Decompositions ($k = 20$)					
Shocks in $\downarrow$	rw	ird	$reer$	ry	cab
rw	0.19	0.05	0.55	0.02	0.04
ird	0.08	0.64	0.02	0.02	0.20
$reer$	0.24	0.06	0.21	0.29	0.16
ry	0.47	0.22	0.16	0.65	0.49
cab	0.02	0.02	0.06	0.02	0.12

To further validate this result, a 6-variable VAR has been formulated by adding US import price as another variable in the VAR, following an over-identified SVAR strategy (Sims-Zha) and impose the restrictions in the matrix below:

$$\begin{bmatrix} 1 & a_{12} & 0 & 0 & a_{15} & 0 \\ 0 & 1 & a_{23} & 0 & a_{25} & 0 \\ a_{31} & a_{32} & 1 & a_{34} & a_{35} & a_{36} \\ a_{41} & 0 & a_{43} & 1 & 0 & a_{46} \\ a_{51} & a_{52} & 0 & 0 & 1 & a_{56} \\ 0 & 0 & a_{63} & a_{64} & 0 & a_{66} \end{bmatrix}$$

Figure 9: IRFs from over-identified SVAR



The technique draws a set of posterior samples from the VAR coefficients and computes impulse responses for each sample. These samples are then summarized to compute MC-based estimates of the responses using the error band methods in Sims and Zha (1999). The confidence bands are drawn by taking draws from the posterior distribution and identifying the shocks. The bands are modelled as the 16 and 84 percentile quantities for the response, which if the distribution is normal, these quantiles would correspond to a one standard deviation band as recommended by Sims and Zha (1999)

6.1 VAR model: essentials

Technical issues involved in VAR analysis like this are illustrated taking examples from Enders (2010) and Patterson (2000) in this section.

Consider first order structural VAR of the following **primitive form**:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1124}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1125}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$. Path of y_t is affected by both contemporaneous and lagged effects of z_t as measured by b_{12} and γ_{12} and its own past values as measured by γ_{11}. Similarly the path of z_t is affected by both contemporaneous and lagged effects of y_t as measured by b_{21} and γ_{21} and its own past values as measured by γ_{22}.

This system can be written in reduced form as

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1126}$$

$$BX_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t \tag{1127}$$

where
$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}, \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}, \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}, X_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix}$
This is transformed in VAR in **standard form** as:

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1128}$$

$A_0 = B^{-1}\Gamma_0; A_1 = B^{-1}\Gamma_1; e_t = B^{-1}\varepsilon_t$
writing explicitly with new notation

$$y_t = a_{10} + a_{12}y_{t-1} + a_{12}z_{t-1} + e_{1t} \tag{1129}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \tag{1130}$$

Errors in the reduced form are composite of errors in the primitive form

$$e_t = B^{-1}\varepsilon_t = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}^{-1} \varepsilon_t = \frac{\begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}'}{1 - b_{21}b_{12}} \varepsilon_t \tag{1131}$$

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} = \frac{1}{1 - b_{21}b_{12}} \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}' \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix} \tag{1132}$$

$$\begin{aligned} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} &= \frac{1}{1 - b_{21}b_{12}} \begin{pmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{pmatrix}' \begin{pmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix} \\ &= \frac{1}{1 - b_{21}b_{12}} \begin{pmatrix} \varepsilon_{yt} - b_{12}\varepsilon_{zt} \\ \varepsilon_{zt} - b_{21}\varepsilon_{yt} \end{pmatrix}' \end{aligned} \tag{1133}$$

$$e_{1t} = \frac{\begin{vmatrix} \epsilon_{yt} & b_{12} \\ \epsilon_{zt} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & b_{12} \\ b_{21} & 1 \end{vmatrix}} = \frac{\epsilon_{yt} \; - \; b_{12}\epsilon_{zt}}{1 - b_{12}b_{21}}; e_{2t} = \frac{\begin{vmatrix} 1 & \epsilon_{yt} \\ b_{21} & \epsilon_{zt} \end{vmatrix}}{\begin{vmatrix} 1 & b_{12} \\ b_{21} & 1 \end{vmatrix}} = \frac{\epsilon_{zt} \; - \; b_{21}\epsilon_{yt}}{1 - b_{12}b_{21}} \tag{1134}$$

Variance-Covariance of Errors in a VAR Mean and variance of composite errors:

$$E(e_{1t}) = E\left[\frac{\epsilon_{yt} \; - \; b_{12}\epsilon_{zt}}{1 - b_{12}b_{21}}\right] = \left[\frac{E(\epsilon_{yt}\;) - \; b_{12}E(\epsilon_{zt})}{1 - b_{12}b_{21}}\right] = 0 \tag{1135}$$

$$E(e_{2t}) = \frac{E(\epsilon_{zt}\;) - \; b_{21}E(\epsilon_{yt})}{1 - b_{12}b_{21}} = 0 \tag{1136}$$

$$Var\left(e_{1t}\right)=E\left[\frac{\epsilon_{yt}\ -\ b_{12}\epsilon_{zt}}{1-b_{12}b_{21}}\right]^2=\frac{\sigma_y^2\ +b_{12}^2\sigma_z^2}{\left(1-b_{12}b_{21}\right)^2} \tag{1137}$$

$$var\left(e_{2t}\right)=E\left[\frac{\epsilon_{zt}\ -\ b_{21}\epsilon_{yt}}{1-b_{12}b_{21}}\right]^2=\frac{\sigma_z^2\ +\ b_{21}^2\sigma_y^2}{\left(1-b_{12}b_{21}\right)^2} \tag{1138}$$

Autocorrelation of Errors in a VAR Thus two shocks are correlated unless contemporaneous effects $b_{12}=0$; $b_{21}=0$.

Autocovariance is time independent.

$$E\left(e_{1t}e_{1t-1}\right)=\frac{E\left[\left(\epsilon_{yt}\ -\ b_{12}\epsilon_{zt}\right)\left(\epsilon_{zt}\ -\ b_{21}\epsilon_{yt}\right)\right]}{\left(1-b_{12}b_{21}\right)^2}=0 \tag{1139}$$

Variance covariance matrix of composite shocks are

$$\sum=\left[\begin{array}{cc} var\left(e_{1t}\right) & cov\left(e_{1t},e_{2t}\right) \\ cov\left(e_{1t},e_{2t}\right) & var\left(e_{2t}\right)\end{array}\right]=\left[\begin{array}{cc}\sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2\end{array}\right] \tag{1140}$$

6.1.1 Stability of VAR

Stability Analysis

$$y_t=a_{10}+a_{12}y_{t-1}+a_{12}z_{t-1}+e_{1t} \tag{1141}$$

$$z_t=a_{20}+a_{21}y_{t-1}+a_{22}z_{t-1}+e_{2t} \tag{1142}$$

Using lag operators

$$y_t=a_{10}+a_{12}Ly_t+a_{12}Lz_t+e_{1t} \tag{1143}$$

$$z_t=a_{20}+a_{21}Ly_t+a_{22}Lz_t+e_{2t} \tag{1144}$$

Collecting terms:

$$\left(1-a_{12}L\right)y_t=a_{10}+a_{12}Lz_t+e_{1t} \tag{1145}$$

$$\left(1-a_{22}L\right)z_t=a_{20}+a_{21}Ly_t+e_{2t} \tag{1146}$$

solve the second equation for z_t and substitute into y_t equation

$$z_t=\frac{a_{20}+a_{21}Ly_t+e_{2t}}{\left(1-a_{22}L\right)} \tag{1147}$$

Putting z_t into y_t equation

$$\left(1-a_{12}L\right)y_t=a_{10}+a_{12}L\left[\frac{a_{20}+a_{21}Ly_t+e_{2t}}{\left(1-a_{22}L\right)}\right]+e_{1t} \tag{1148}$$

Collecting terms:

$$
\begin{aligned}
(1-a_{12}L)(1-a_{22}L)\,y_t &= (a_{10}1-a_{22}L) \\
&\quad +a_{12}a_{20}+\left[a_{12}a_{21}L^2y_t+a_{12}Le_{2t}\right] \\
&\quad +(1-a_{22}L)\,e_{1t}
\end{aligned} \tag{1149}
$$

$$
\begin{aligned}
(1-a_{12}L)(1-a_{22}L)\,y_t-a_{12}a_{21}L^2y_t &= a_{10}\,(1-a_{22}L) \\
&\quad +a_{12}a_{20}+a_{12}Le_{2t} \\
&\quad +(1-a_{22}L)\,e_{1t}
\end{aligned} \tag{1150}
$$

$$
y_t=\frac{a_{10}\,(1-a_{22})+a_{12}a_{20}+a_{12}Le_{2t}+(1-a_{22}L)\,e_{1t}}{(1-a_{12}L)(1-a_{22}L)-a_{12}a_{21}L^2} \tag{1151}
$$

$$
z_t=\frac{a_{10}\,(1-a_{11})+a_{21}a_{10}+a_{21}Le_{2t-1}+(1-a_{11}L)\,e_{2t}}{(1-a_{12}L)(1-a_{22}L)-a_{12}a_{21}L^2} \tag{1152}
$$

Convergence requires the roots of the polynomial (inverse characteristic equation), $(1-a_{12}L)(1-a_{22}L)-a_{12}a_{21}L^2$,lie outside the unit circle (or equivalently roots of the characteristic equation lie inside the unit circle. Since both y_t and z_t equations have same polynomial $(1-a_{12}L)(1-a_{22}L)-a_{12}a_{21}L^2$ time series of both variable are similar. These depend on whether the roots are real, distinct or complex.

Convergence requires the roots of the polynomial (inverse characteristic equation),
$(1-a_{12}L)(1-a_{22}L)-a_{12}a_{21}L^2=0$
$1-a_{12}L-a_{22}L+a_{12}a_{22}L^2-a_{12}a_{21}L^2=0$
Putting $L=\frac{1}{\lambda}$
$\lambda^2-(a_{12}+a_{22})\,\lambda+(a_{12}a_{22}-a_{12}a_{21})=0$
Quadratic roots

$$
\lambda_1,\lambda_2=\frac{(a_{12}+a_{22})\pm\sqrt{(a_{12}+a_{22})^2-4\,(a_{12}a_{22}-a_{12}a_{21})}}{2} \tag{1153}
$$

If roots λ_1,λ_2 lie within the unit circle each variable is stationary and it can not be cointegrated for order 1, C(1,1). If both roots λ_1,λ_2 lie outside the unit circles and processes are explosive and cannot be cointegrated of order 1,y_t and z_t are explosive, if and then then two variables evolve without any long run relationship; y_t and z_t have cointegration of order 1, C(1,1) only if one of the roots λ_1,λ_2 is unity and another is less than unity (see more on this topic in Elders Chapter 6.2)

VAR Experiments: generate 100 random values of e_{1t} and e_{2t}.
a) $a_{10}=a_{20}=0$; $a_{11}=a_{22}=0.7$; $a_{12}=a_{21}=0.2$.
b) $a_{10}=a_{20}=0$; $a_{11}=a_{22}=0.5$; $a_{12}=a_{21}=-0.2$.
c) $a_{10}=a_{20}=0$; $a_{11}=a_{22}=0.5$; $a_{12}=a_{21}=0.5$.
d) $a_{10}=0.5; a_{20}=0$; $a_{11}=a_{22}=0.7$; $a_{12}=a_{21}=0.2$.
A VAR(p) process on X variables is stationary when

- when all variables have constant mean $E(x)=\mu$

- when its variance is finite, $VAR(X_t) < \infty$
- Covariance is time independent, $cov(X_t X_{t-1}) = E\left[(X_t - \mu)((X_{t+2} - \mu)'\right] = \Gamma_k \quad for \;\; \text{all } t$

VAR: Example 2

$$y_{1t} = \nu_1 + \theta_{12} y_{1t-1} + \theta_{12} y_{2t-1} + e_{1t} \tag{1154}$$

$$y_{2t} = \nu_2 + \theta_{21} y_{1t-1} + \theta_{22} y_{2t-1} + e_{2t} \tag{1155}$$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1156}$$

Polynomial method:
$\det\left[I - \theta_1 Z - \theta_2 Z^2 - ....\theta_p Z^p\right] = 0$
$\det\left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix} Z\right\} = det \begin{bmatrix} 1-\theta_{11}z & -\theta_{12}z \\ -\theta_{21}z & 1-\theta_{22}z \end{bmatrix} \Rightarrow \begin{vmatrix} 1-\theta_{11}z & -\theta_{12}z \\ -\theta_{21}z & 1-\theta_{22}z \end{vmatrix} = 0$
$(1-\theta_{11}z)(1-\theta_{22}z) - \theta_{12}\theta_{21}z^2 = 0$
z_1 and z_1 lie outside of the unit circle.
z_1 and z_1 lie outside of the unit circle.
if $\theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = \begin{bmatrix} 0.008 & 0.461 \\ 0.232 & 0.297 \end{bmatrix}$
Characteristic root method
$A = \begin{bmatrix} 0.008 & 0.461 \\ 0.232 & 0.297 \end{bmatrix} \quad |A - \lambda I| = \begin{vmatrix} 0.008-\lambda & 0.461 \\ 0.232 & 0.297-\lambda \end{vmatrix} = 0$ from this $\lambda_1 = -0.2075$ and $\lambda_2 = 0.521$
Identification of the order of the VAR

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + ... + A_2 X_{t-p} + e_t \tag{1157}$$

whereA_0 is $N \times 1$,matrix A_1 $N \times N$ matrix, $e_t \sim N \times 1$, $X \sim N \times 1$ vector. There are $n^2 p + n$ number of parameters. Normally do not difference variables as it throws away information. VAR in levels is considered better than VAR in the first differences. For instance, consider again

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t} \tag{1158}$$

$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t} \tag{1159}$$

is this system identified? Can the structural parameters in its primitive form

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \epsilon_{yt} \tag{1160}$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1161}$$

6.1.2 Identification of the order of the VAR

Can ten parameters b_{10}, b_{20}, b_{12}, b_{21}, γ_{11} , γ_{21},, γ_{12} , γ_{22} and $var\left(\epsilon_{yt}\right)$ and $var\left(\epsilon_{zt}\right)$ be retrieved from the nine reduced for parameters a_{10}, a_{20}, a_{12}, a_{21}, $a_{11}, a_{22}, var\left(e_{1t}\right)$ and $var\left(e_{2t}\right)$ and $covar\left(e_{1t}e_{1t}\right)$? Nine parameters cannot give unique values for ten parameters. It needs one more equation. This can be made possible by imposing one restriction such as $b_{21} = 0$ meaning that there is no contemporaneous effect from y_t to z_t .This restricted system can be written as

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1162}$$

$$z_t = b_{20t} + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1163}$$

with this restriction $B^{-1} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1164}$$

$$\begin{aligned} \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \\ & \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \end{aligned} \tag{1165}$$

$$\begin{aligned} \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} b_{10} - b_{10}b_{12} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} - b_{12}\gamma_{21} & \gamma_{12} - b_{12}\gamma_{22} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \\ & \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & \epsilon_{yt} - b_{12}\epsilon_{zt} \\ 0 & \epsilon_{zt} \end{bmatrix} \end{aligned} \tag{1166}$$

Now the estimated coefficients of the reduced form equations can be identified to the coefficients of the restricted structural equations as

$$a_{10} = b_{10} - b_{10}b_{12}; \quad a_{11} = \gamma_{11} - b_{12}\gamma_{21} \tag{1167}$$

$$a_{12} = \gamma_{12} - b_{12}\gamma_{22}; a_{20} = b_{20}; \quad a_{21} = \gamma_{21}; a_{22} = \gamma_{12} \tag{1168}$$

$$var\left(e_1\right) = \sigma_y^2 + b_{12}^2\sigma_z^2 \; ; var\left(e_1\right) = \sigma_z^2 \tag{1169}$$

$$covar\left(e_1, e_2\right) = -b_{12}^2\sigma_z^2 \tag{1170}$$

- Now the structural residuals ϵ_{yt} and ϵ_{zt} can be estimated. While y_t.is affected by both ϵ_{yt} and ϵ_{zt} shocks z_t is only affected by ϵ_{zt} .
- Given these parameters both ϵ_{yt} and ϵ_{zt} sequences can be retrieved from the structural parameters and sequences of e_{1t} and e_{2t} .
- Triangular restriction like this are called Cholesky decomposition.

Over Identification Restrictions

If $\gamma_{21} = 0$ is combined with $b_{22} = 0$ then z_t is free from any impacts from y_t.Testing $a_{21} = 0$ is equivalent to testing $\gamma_{21} = 0$

Overidentified. restrictions:

Consider $\gamma_{12} = \gamma_{21} = 0$

$$y_t = b_{10} + \gamma_{11} y_{t-1} + b_{12} z_t + \epsilon_{yt} \tag{1171}$$

$$z_t = b_{20} + b_{21} y_t + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1172}$$

By direct substitution of y_t to z_t

$$y_t = b_{10} + \gamma_{11} y_{t-1} + b_{12} \left(b_{10} + \gamma_{11} y_{t-1} + b_{12} z_t + \epsilon_{yt} \right) + \epsilon_{yt} \tag{1173}$$

$$z_t = b_{20} + b_{21} \left(b_{20} + b_{21} y_t + \gamma_{22} z_{t-1} + \epsilon_{zt} \right) + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1174}$$

Over Identification Restrictions

$$y_t = a_{10} + a_{11} y_{t-1} + a_{12} z_{t-1} + e_{1t} \tag{1175}$$

$$z_t = a_{20} + a_{21} y_{t-1} + a_{22} z_{t-1} + e_{2t} \tag{1176}$$

There are nine parameters in the reduced form but only eight in the structural equations. Therefore this VAR system is over identified. More than one value of structural parameter are permissible from the reduced form estimates.

$$a_{10} = \frac{b_{10} + b_{10} b_{12}}{(1 - b_{12}\gamma_{21})}; \quad a_{11} = \frac{\gamma_{11}}{(1 - b_{12}\gamma_{21})}; a_{12} = \frac{b_{12}\gamma_{22}}{(1 - b_{12}\gamma_{21})}; \tag{1177}$$

$$a_{20} = \frac{b_{20} - b_{20} b_{21}}{(1 - b_{12}\gamma_{21})}; \quad a_{21} = \frac{\gamma_{11}\gamma_{21}}{(1 - b_{12}\gamma_{21})}; a_{22} = \frac{\gamma_{22}}{(1 - b_{12}\gamma_{21})} \tag{1178}$$

$$var(e_1) = \frac{\sigma_y^2 + b_{12}^2 \sigma_z^2}{(1 - b_{12}\gamma_{21})^2}; \ var(e_1) = \frac{\sigma_y^2 + b_{12}^2 \sigma_z^2}{(1 - b_{12}\gamma_{21})^2}; \tag{1179}$$

$$covar(e_1, e_2) = \frac{b_{21}\sigma_y^2 - b_{12}\sigma_z^2}{(1 - b_{12}\gamma_{21})^2} \tag{1180}$$

6.1.3 Impulse Response Analysis

$$y_t = a_{10} + a_{12} L y_t + a_{12} L z_t + e_{1t} \tag{1181}$$

$$z_t = a_{20} + a_{21} L y_t + a_{22} L z_t + e_{2t} \tag{1182}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1183}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y} \\ \overline{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \tag{1184}$$

Impulse Response Analysis

Given

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \left[\frac{1}{1-b_{12}b_{21}}\right] \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1185}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y} \\ \overline{z} \end{bmatrix} + \sum_{i=0}^{\infty} \left[\frac{1}{1-b_{12}b_{21}}\right] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1186}$$

Introducing more simplifying assumptions:

$\phi_i = \left[\frac{A^i}{1-b_{12}b_{21}}\right] \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix}$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y} \\ \overline{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \epsilon_{yt-i} \\ \epsilon_{zt-i} \end{bmatrix} \tag{1187}$$

$\phi_{11}(i), \phi_{12}(i), \phi_{12}(i)$ and $\phi_{12}(i)$ are impulse response functions.

6.1.4 Variance Decomposition

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i(i)\epsilon_{t-i} \tag{1188}$$

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1189}$$

from successive iteration this reduces to

$$E_t X_{t+n} = \left(I + A_1 + A_1^2 + A_1^3 + ... + A_1^{n-1}\right) A_0 + A_1^n X_t + e_t \tag{1190}$$

Forecast error is given by

$$\left(e_{t+n} + A_1 e_{t+n-1} + A_1^2 e_{t+n-2} + ... + A_1^{n-1} e_{t+1}\right) A_0 + A_1^n X_t \tag{1191}$$

$$X_{t+n} - E_t X_{t+n} = \sum_{i=0}^{n-1} \phi_i(i)\epsilon_{t+n-i} \tag{1192}$$

Taking only one equation

$$\begin{aligned} y_{t+n} - E_t y_{t+n} = {} & \phi_{11}(0)\epsilon_{yt+n} + \phi_{11}(1)\epsilon_{yt+n-1} + ... + \phi_{11}(n-1)\epsilon_{yt+1} \\ & + \phi_{12}(0)\epsilon_{zt+n} + \phi_{12}(1)\epsilon_{zt+n-1} + ... + \phi_{12}(n-1)\epsilon_{zt+1} \end{aligned}$$

Variance of n-step ahead forecast error is

$$\begin{aligned} \sigma(n)_y^2 &= \sigma_y^2 \left[\phi_{11}(0) + \phi_{11}(1) + ... + \phi_{11}(n-1)\right] \\ &\quad + \sigma_z^2 \left[\phi_{12}(0) + \phi_{12}(1) + ... + \phi_{12}(n-1)\right] \end{aligned} \tag{1193}$$

Variance decomposition in terms of variances of shocks ϵ_{yt} and ϵ_{zt}.

$$\begin{aligned} 1 &= \frac{\sigma_y^2 \left[\phi_{11}(0) + \phi_{11}(1) + ... + \phi_{11}(n-1)\right]}{\sigma(n)_y^2} \\ &\quad + \frac{\sigma_z^2 \left[\phi_{12}(0) + \phi_{12}(1) + ... + \phi_{12}(n-1)\right]}{\sigma(n)_y^2} \end{aligned} \tag{1194}$$

Thus the variance decomposition is finding the proportion of variance explained by variables' its own shock(ϵ_{yt}) versus the variance explained by shock of the other variable (ϵ_{yt}).

6.1.5 Lag Selection

How many lags are appropriate in a VAR model? This essentially depends on tests. Estimate the model with the OLS if all variables have same number of lags or with seemingly unrelated regression (SUR) method if lags differ for different variables. For instance if X contains monthly data first estimate a model with 12 lags. If there are 5 variables this means estimating $(np+n) = 5 \times (12+1) = 65$ coefficient per equation or $\left(n^2 p + n\right) = 25 \times 12 + 5 = 305$ coefficients for the system. Variance-Covariance of error $\sum$ is of dimension $N \times N$. F test can be used to find whether a different lag is significant. For this compute the F-ratio $\sum_8$ to $\sum_{12}$ for each equation. Likelihood ratio test is suggested for the system wise estimation. First estimated the unrestricted model get the $\sum_{12}$. Then estimate the restricted model say with eight lags $\sum_8$ .Log-likelihood ratio is defined as

$$L = T\left\{\log\left|\sum_{8}\right| - \log\left|\sum_{12}\right|\right\} \tag{1195}$$

The degrees of freedom for this equal the number of restriction. This is $4n$ for each equation and $4n^2 = 100$ for the entire system. Sims(1986) modifies the likelihood function as $L = (T-c)\left\{\log\left|\sum_8\right| - \log\left|\sum_{12}\right|\right\}$ where c is the number of parameters estimated in the unrestricted system, $c = 12n+1$ with 12 quarters and 5 variables.

For generic models

$$\begin{bmatrix} X_{1t} \\ X_{2t} \\ . \\ X_{nt} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ . \\ A_{n0} \end{bmatrix}$$
$$+\begin{bmatrix} A_{11}(L) & A_{12}(L) & . & A_{1n}(L) \\ A_{21}(L) & A_{22}(L) & . & A_{2n}(L) \\ . & . & . & . \\ A_{n1}(L) & A_{n2}(L) & . & A_{nn}(L) \end{bmatrix}\begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ . \\ X_{nt-1} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \\ . \\ e_{nt} \end{bmatrix} \tag{1196}$$

Likelihood ratio test tor generic restricted and unrestricted systems :

$$L = (T-c)\left\{\log\left|\sum_{r}\right| - \log\left|\sum_{u}\right|\right\} \tag{1197}$$

Minimum of the Akaike Information Criteria (AIC) and SBC criteria are also used to determine the lag structure.

$$AIC = T\log\left|\sum\right| + 2N \tag{1198}$$

$$SBC = T\log\left|\sum\right| + 2\log T \tag{1199}$$

Estimate VAR with different lags and select the model with minimum AIC or SBC numbers. Granger causality (block causality) test can be used to find the direction of causality from one to another variable. This depends on the validity of $a_{ij}(1), a_{ij}(2)$..parameters in $A_{ij}(L)$.

6.1.6 Tutorial 3: VAR and cointegration analysis

Q1. Consider a structural VAR model between y_t and z_t as following:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1200}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1201}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$.

a. Derive the reduced form of this VAR model and suggest how the maximum likehoold estimator (MLE) could be applied to estimate the parameters in it.

b. How would one determine stability of a VAR system like this? Provide analytical solutions using the roots of the quadratic function.

c. How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified?

d. Write impulse response functions for these two equations and indicate how can one perform an impulse response analysis with them?

e. What is the meaning of variance decomposition in a VAR model like this?

f. Why is the Bayesian VAR becoming more popular than a classical VAR as given above in recent years?

Q2. Consider a vector error correction model (VECM) of the form

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1202}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_3 + ... + \Pi_i \quad \text{for } i = 1..k \quad \Gamma_k$ gives the long run solution and $\varepsilon_t \sim N(0, \sigma_y^2)$.

a. What is the meaning of cointegration and why should there exist at least one cointegrating vector in this equation?

b. Discuss how likelihood ratio tests are employed to determine the optimal lag.

c. Explain how the canonical correlations provide eigenvalues and eigen vectors that are useful in determining the rank of the cointegrating vector.

d. Discuss a procedure for trace and max-eigenvalue tests for cointegration.

6.1.7 Structural VAR

VAR is a-theoretic. VAR modelers put restriction on coefficients base on their hunch on economic theory. These restrictions can be in parameters, variance covariance matrices or symmetry.

Example of restrictions on parameters:

Coefficient restriction (macro model estimated in Sims(1986) as reported by Enders(1995).

VAR of gdp(y), interest rate (r), money supply (m), price level (p), unemployment (u) and investment (i).

$$\begin{bmatrix} 1 & -71.1 & 0 & 0 & 0 & 0 \\ 0.008 & 1 & 0.283 & 0.224 & 0 & 0 \\ 0.00135 & 0 & 1 & 0 & 0 & -0.1324 \\ 0.001 & 0 & -0.045 & 1 & 0 & 0.0086 \\ 0.116 & 0 & 20.1 & 8.89 & 1 & 1.48 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ m \\ y \\ p \\ u \\ i \end{bmatrix} = \begin{bmatrix} \varepsilon_{rt} \\ \varepsilon_{mt} \\ \varepsilon_{yt} \\ \varepsilon_{pt} \\ \varepsilon_{ut} \\ \varepsilon_{it} \end{bmatrix} \tag{1203}$$

Take structural model in line with Sim (1986) and Bernanke(1986) system

$$\begin{bmatrix} 1 & b_{12} & . & b_{1n} \\ b_{21} & 1 & . & b_{2n} \\ . & . & . & . \\ b_{n1} & b_{n2} & . & b_{nn} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ . \\ X_{nt} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ . \\ b_{n0} \end{bmatrix}$$

$$+ \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdot & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdot & \gamma_{2n} \\ . & . & . & . \\ \gamma_{n1} & \gamma_{n2} & \cdot & \gamma_{nn} \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ . \\ X_{nt-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ . \\ \varepsilon_{nt} \end{bmatrix} \tag{1204}$$

or compactly with ($A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$, $e_t = B^{-1}\varepsilon_t$)

$$X_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 X_{t-1} + B^{-1}\varepsilon_t \tag{1205}$$

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1206}$$

Reduced form is estimated with the available data; then structural shocks are retrieved using $e_t = B^{-1}\varepsilon_t$ This requires estimation of the variance covariance matrix of the error term

$$\sum = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & . & \sigma_{2n} \\ . & . & . & . \\ \sigma_{n1} & \sigma_{n2} & . & \sigma_{nn} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & . & \sigma_{2n} \\ . & . & . & . \\ \sigma_{n1} & \sigma_{n2} & . & \sigma_n^2 \end{bmatrix} \tag{1207}$$

$\sigma_{ij} = \frac{1}{T}\sum\limits_{t=1}^{T} e_{ij}e_{ij}^{'}$

Consider the same structural model considered earlier:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1208}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1209}$$

$$y_t = a_{10} + a_{12}y_{t-1} + a_{12}z_{t-1} + e_{1t} \tag{1210}$$

$$z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \tag{1211}$$

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1-b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix} \tag{1212}$$

Variance restrictions
$e_t = B^{-1}\varepsilon_t$
$e_{1t} = \epsilon_{1t}$
$e_{2t} = c_{21}e_{1t} + \epsilon_{2t}$
$e_{3t} = c_{31}e_{1t} + c_{32}e_{2t} + \epsilon_{3t}$

Thus ϵ_{1t},$\epsilon_{2t}$and ϵ_{3t} can be identified from estimates of e_{1t},$e_{2t}$and e_{3t} and their variance covariance matrices such as (two variable case) $\sum = e_t e_t' = \begin{bmatrix} e_{1t}^2 & e_{1t}e_{2t} \\ e_{2t}e_{1t} & e_{2t}^2 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}$; $\sum = \frac{1}{T}\sum\limits_{t=1}^{T} e_t e_t'$

$\sum_\varepsilon = \begin{bmatrix} var(\epsilon_{1t}) & 0 \\ 0 & var(\epsilon_{2t}) \end{bmatrix}$; $\sum_\varepsilon = \frac{1}{T}\sum\limits_{t=1}^{T} \epsilon_t \epsilon_t'$

Structural VAR

From $e_t = B^{-1}\varepsilon_t$ it is possible to find the links between $\sum_\varepsilon$ and $\sum$.

$\sum_\varepsilon = \frac{1}{T}\sum\limits_{t=1}^{T} \epsilon_t \epsilon_t' = \frac{1}{T}\sum\limits_{t=1}^{T} (Be_t)(e_t'B')$

$\sum_\varepsilon = B\sum B'$

$\sum_\varepsilon = \begin{bmatrix} var(\epsilon_{1t}) & 0 \\ 0 & var(\epsilon_{2t}) \end{bmatrix} = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0.4 \\ 0.4 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & b_{21} \\ b_{12} & 1 \end{bmatrix}$

By expanding these numbers

$\begin{bmatrix} var(\epsilon_{1t}) & 0 \\ 0 & var(\epsilon_{2t}) \end{bmatrix} = \begin{bmatrix} 0.5 + 0.4b_{12} + b_{12}(0.4 + 0.5b_{12}) & 0.5b_{21} + 0.4 + b_{12}(0.4b_{21} + 0.5) \\ b_{21}(0.5 + 0.4b_{12}) + 0.4 + 0.5b_{12} & b_{21}(0.5b_{21} + 0.4) + (0.4b_{21} + 0.5) \end{bmatrix}$

Structural VAR

$var(\epsilon_{1t}) = 0.5 + 0.4b_{12} + b_{12}(0.4 + 0.5b_{12}) = 0.5 + 0.8b_{12} + 0.5b_{12}^2$
$0 = 0.5b_{21} + 0.4 + b_{12}(0.4b_{21} + 0.5)$
$0 = b_{21}(0.5 + 0.4b_{12}) + 0.4 + 0.5b_{12} = 0.5b_{21} + 0.4b_{12}b_{21} + 0.4 + 0.5b_{12}$
$var(\epsilon_{2t}) = b_{21}(0.5b_{21} + 0.4) + (0.4b_{21} + 0.5) = 0.5b_{21}^2 + 0.8b_{21} + 0.5$

when restriction $b_{12} = 0$ valid, then $var(\epsilon_{1t}) = 0.5; b_{21} = -0.8$; and $var(\epsilon_{2t}) = 0.18$;

Now the structural shocks ϵ_{1t} and ϵ_{2t}can be retrieved from above estimations.

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1-b_{12}b_{21}} \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{1213}$$

$$\epsilon_{1t} = e_{1t}\ and \epsilon_{2t} = -0.8e_{1t} + e_{2t} \tag{1214}$$

$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}$ is identified as $\begin{bmatrix} 1 & 0 \\ -0.8 & 1 \end{bmatrix}$
when restriction $b_{21} = 0$ valid, then $var\left(\epsilon_{1t}\right) = 0.18; b_{12} = -0.8;$ and $var\left(\epsilon_{2t}\right) = 0.5;$
$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix}$ is identified as $\begin{bmatrix} 1 & -0.8 \\ 0 & 1 \end{bmatrix}$
Thus the ordering of equations is important in VAR.
The structural errors can now be retrieved from the
Structural VAR: Restrictions
Symmetry Restrictions (Cholesky decomposition)
$b_{12} = b_{13} = b_{14} = ... = b_{1n} = 0$
$b_{23} = b_{24} = ... = b_{2n} = 0$
$b_{24} = ... = b_{2n} = 0$
....
$b_{nn} = 0$

There are $\frac{\left(n^2+n\right)}{2}$ distinct elements in the structural coefficient of $N \times N$ VAR and it requires $n^2 - \frac{\left(n^2+n\right)}{2} = \frac{\left(n^2-n\right)}{2}$ restrictions.

Stability of the VAR system

$$X_t = A_0 + A_1 X_{t-1} + \varepsilon_t \tag{1215}$$

for stability check the homegenous part of the solution

$$X_t = A_1 X_{t-1} \tag{1216}$$

use the undetermined coefficient to solve this problem.

$$X_{i,t} = C_i \lambda^t \tag{1217}$$

Eigen Values and Stability of VAR System Insert these in the extended VAR

$$C_1\lambda^t = a_{11} C_1 \lambda^{t-1} + a_{12} C_2 \lambda^{t-1} + a_{13} C_3 \lambda^{t-1} + ... + a_{1,n} C_n \lambda^{t-1} \tag{1218}$$

$$C_2\lambda^t = a_{21} C_1 \lambda^{t-1} + a_{22} C_2 \lambda^{t-1} + a_{23} C_3 \lambda^{t-1} + ... + a_{2,n} C_n \lambda^{t-1} \tag{1219}$$

$$.. \tag{1220}$$

$$C_n\lambda^t = a_{n1} C_1 \lambda^{t-1} + a_{n,2} C_2 \lambda^{t-1} + a_{n,3} C_3 \lambda^{t-1} + ... + a_{n,n} C_n \lambda^{t-1} \tag{1221}$$

For homogenous case

$$C_1\left(a_{11} - \lambda\right) + a_{12} C_2 + a_{13} C_3 + ... + a_{1,n} C_n = 0 \tag{1222}$$

$$C_2\lambda^t = a_{21} C_1 + C_2\left(a_{11} - \lambda\right) + a_{23} C_3 + ... + a_{2,n} C_n = 0 \tag{1223}$$

$$.. \tag{1224}$$

$$C_n\lambda^t = a_{n1}C_1 + a_{n,2}C_2 + a_{n,3}C_3 + ... + C_n\left(a_{nn} - \lambda\right) = 0 \tag{1225}$$

$$\begin{bmatrix} (a_{11} - \lambda) & a_{12} & a_{13} & a_{nn} \\ a_{21} & (a_{11} - \lambda) & . & a_{2,n} \\ . & . & . & . \\ a_{n1} & a_{n,2} & . & (a_{nn} - \lambda) \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ . \\ C_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{1226}$$

Non trivial solution requires

$$\begin{vmatrix} (a_{11} - \lambda) & a_{12} & a_{13} & a_{nn} \\ a_{21} & (a_{11} - \lambda) & . & a_{2,n} \\ . & . & . & . \\ a_{n1} & a_{n,2} & . & (a_{nn} - \lambda) \end{vmatrix} = 0 \tag{1227}$$

The determinant will be n order polynomial and n values of $\lambda_1, \lambda_2, ..., \lambda_n$.Necessary and sufficient condition for stability is that all eigen values (characteristic roots) lie within the unit circle.

A is $N \times N$ matrix. Then the determinants of $|A| = \prod\limits_{i=1}^{N} \lambda_i$ for i $=\lambda_1, \lambda_2, ..., \lambda_n$

$|A - \lambda I| = 0$ implies

$(a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) .... (a_{nn} - \lambda) = 0$

$\lambda^n + b_1\lambda^{n-1} + b_2\lambda^{n-2} + ... + b_{n-1}\lambda + b_n = 0$

by factor rule of the polynomial $\prod\limits_{i=1}^{N} \lambda_i = (-1)^n b_n = |A| = \lambda_1\lambda_2\lambda_3.....\lambda_{n-1}\lambda_n$

Rank of A equals the number of non-zero characteristic roots. If $|A| \neq 0$ then non of $\lambda_i = 0$. A has full rank. If rank of A is zero then each element of A must be zero and $\lambda_1 = \lambda_2 = \lambda_3 = ... = \lambda_{n-1} = \lambda_n = 0$. In intermediate case rank of A is between 0 and N.

6.1.8 Blanchard-Qua decomposition

Blanchard-Qua (1989) Decomposition of temporary and permanent Components of a I(1) Process (Enders (12))

Let $y_t \sim I(1)$ be output and $z_t \sim I(0)$ be unemployment ; $\epsilon_{1,t}$ is the demand shock $\epsilon_{2,t}$ the supply shock

$$\Delta y_t = \sum_{k=0}^{\infty} c_{11}(k)\epsilon_{1,t-k} + \sum_{k=1}^{\infty} c_{12}(k)\epsilon_{2,t-k} \tag{1228}$$

$$z_t = \sum_{k=0}^{\infty} c_{21}(k)\epsilon_{1,t-k} + \sum_{k=1}^{\infty} c_{22}(k)\epsilon_{2,t-k} \tag{1229}$$

$$\begin{bmatrix} \Delta y_t \\ z_t \end{bmatrix} = \begin{bmatrix} c_{11}(L) & c_{12}(L) \\ c_{21}(L) & c_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{1230}$$

$c_{11}(L)$ impulse of demand shock on log GDP y_t If the demand shock does not have any long run impact on output $\sum\limits_{k=0}^{\infty} c_{11}(k)\epsilon_{1,t-k} = 0$

$\sum_\epsilon = \begin{bmatrix} var\left(\epsilon_{1t}\right) & cov\left(\epsilon_{1t}, \epsilon_{2t}\right) \\ cov\left(\epsilon_{1t}, \epsilon_{2t}\right) & var\left(\epsilon_{2t}\right) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{bmatrix} \Delta y_t \\ z_t \end{bmatrix} = \begin{bmatrix} A_{11}(L) & A_{12}(L) \\ A_{21}(L) & A_{22}(L) \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1231}$$

Important point in the BQ decomposition is that ϵ_{1t} and ϵ_{2t} are not directly observable

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} c_{11}(0) & c_{12}(0) \\ c_{21}(0) & c_{22}(0) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \tag{1232}$$

$$e_{1t} = c_{11}(0)\epsilon_{1t} + c_{12}(0)\epsilon_{2t} \tag{1233}$$

$$e_{2t} = c_{21}(0)\epsilon_{1t} + c_{22}(0)\epsilon_{2t} \tag{1234}$$

But they could be computed $c_{11}(0), c_{12}(0), c_{21}(0), c_{22}(0)$ were known using the residuals estimated from the reduced form:

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} = \begin{bmatrix} c_{11}(0) & c_{12}(0) \\ c_{21}(0) & c_{22}(0) \end{bmatrix}^{-1} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1235}$$

Restriction 1

From assumption $E\left(\epsilon_{1,t}\epsilon_{2,t}\right) = 0 \quad var\left(\epsilon_{1,t}\right) = 1; \quad var\left(\epsilon_{2,t}\right) = 1$

$$var\left(e_{1t}\right) = c_{11}(0)^2 + c_{12}(0)^2 \tag{1236}$$

Restriction 2

$$var\left(e_{2t}\right) = c_{21}(0)^2 + c_{22}(0)^2 \tag{1237}$$

Restriction 3

$E\left(e_{1,t}e_{2,t}\right) = E\left[\left\{c_{11}(0)\epsilon_{1t} + c_{12}(0)\epsilon_{2t}\right\}\left\{c_{21}(0)\epsilon_{1t} + c_{22}(0)\epsilon_{2t}\right\}\right]$

$$E\left(e_{1,t}e_{2,t}\right) = \left[c_{11}(0)c_{21}(0) + c_{12}(0)c_{22}(0)\right] \tag{1238}$$

Workout for this assumption

$$X_t = A(L)LX_t + e_t \tag{1239}$$

$$\left(I - A(L)L\right)X_t = e_t \tag{1240}$$

$$X_t = \left[I - A(L)L\right]^{-1} e_t \tag{1241}$$

$$\begin{bmatrix} \Delta y_t \\ z_t \end{bmatrix} = \frac{1}{D} \begin{bmatrix} 1 - A_{22}(L)L & A_{12}(L)L \\ A_{21}(L) & 1 - A_{11}(L)L \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1242}$$

$$\begin{bmatrix} \Delta y_t \\ z_t \end{bmatrix} = \frac{1}{D} \begin{bmatrix} 1 - \sum a_{22}(k)L^{k+1} & \sum a_{12}(k)L^{k+1} \\ \sum a_{21}L^{k+1} & 1 - \sum a_{11}(k)L^{k+1} \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1243}$$

$$\Delta y_t = \frac{1}{D}\left\{\left[1 - \sum_{k=0}^{\infty} a_{22}(k)L^{k+1}\right] e_{1t} + \sum_{k=0}^{\infty} a_{12}(k)L^{k+1} e_{2t}\right\} \tag{1244}$$

$$\left[1-\sum_{k=0}^{\infty}a_{22}(k)L^{k+1}\right]c_{11}(0)\epsilon_{1t}+\sum_{k=0}^{\infty}a_{12}(k)L^{k+1}c_{21}(0)\epsilon_{2t}=0 \tag{1245}$$

Restriction 4

It comes from $\sum_{k=0}^{\infty}c_{11}(k)\epsilon_{1,t-k}=0$

$$\left[1-\sum_{k=0}^{\infty}a_{22}(k)\right]c_{11}(0)+\sum_{k=0}^{\infty}a_{12}(k)c_{21}(0)=0 \tag{1246}$$

Thus there are four equations that can be used to identify four parameters $c_{11}(0)$, $c_{21}(0)$, $c_{12}(0)$ and $c_{22}(0)$.

6.1.9 Long run and short run multipliers

Let consumption (C_t) function function of income (X_t) as:

$$C_t=\beta_0+\beta_1X_t+\beta_2X_{t-1}+\beta_3X_{t-1}+....+\beta_kX_{t-k}+u_t \tag{1247}$$

Here short run multiplier: β_1

Intermediate multiplier: $\beta_1+\beta_2+\beta_3$

Long run multiplier : $\beta=\sum_{k=0}^{\infty}(\beta_1+\beta_2+\beta_3+...+\beta_k)$

By Koyck procedure $\beta_2=\lambda\beta_1;\beta_3=\lambda^2\beta_1;\beta_k=\lambda^k\beta_1$; where $0<\lambda<1$

$$C_t=\beta_0+\beta_1X_t+\lambda\beta_1X_{t-1}+\lambda^2\beta_1X_{t-1}+....+\lambda^k\beta_1X_{t-k}+u_t \tag{1248}$$

$$\begin{aligned}\lambda C_{t-1} &= \lambda\beta_0+\lambda\beta_1X_{t-1}+\lambda^2\beta_1X_{t-2}+\lambda^3\beta_1X_{t-3}+....+\lambda^{k+1}\beta_1X_{t-k-1}\\ &\quad +u_{t-1}\end{aligned} \tag{1249}$$

By taking these two differences

$$C_t-\lambda C_{t-1}=\beta_0-\lambda\beta_0+\beta_1X_t-\lambda^{k+1}\beta_1X_{t-k-1}+u_t-u_{t-1} \tag{1250}$$

$$C_t=\beta_0-\lambda\beta_0+\beta_1X_t+\lambda C_{t-1}+e_t \tag{1251}$$

$e_t=u_t-u_{t-1}$

Use minimise AIC and SBC criteria while choosing the optimal lag.

Example 2

$$Y_t = \delta + \theta Y_{t-1} + \phi_0 X_t + \phi_1 X_{t-1} + u_t \tag{1252}$$

short impact multiplier : $\frac{\partial Y_t}{\partial X_t} = \phi_0$

Lag it further

$$Y_{t+1} = \delta + \theta Y_t + \phi_0 X_{t+1} + \phi_1 X_t + u_{t+1} \tag{1253}$$

$\frac{\partial Y_{t+1}}{\partial X_t} = \frac{\theta \partial Y_t}{\partial X_t} = \theta\phi_0 + \phi_1$

Continuing this process

$$Y_{t+2} = \delta + \theta Y_{t+1} + \phi_0 X_{t+2} + \phi_1 X_{t+1} + u_{t+2} \tag{1254}$$

$\frac{\partial Y_{t+2}}{\partial X_t} = \frac{\theta \partial Y_{t+1}}{\partial X_t} = \frac{\theta \partial Y_t}{\partial X_t} = \theta\left(\theta\phi_0 + \phi_1\right)$

If this process continues, the long run multiplier is given by $\phi_0 + (\theta\phi_0 + \phi_1) + \theta\left(\theta\phi_0 + \phi_1\right) + ... + = \phi_0 + \left(1 + \theta + \theta^2 + ...\right)\left(\theta\phi_0 + \phi_1\right)$ Since $\theta < 1$

Long run and short run multipliers

$$\phi_0 + (\theta\phi_0 + \phi_1) + \theta\left(\theta\phi_0 + \phi_1\right) + ... + = \frac{\phi_0 + \phi_1}{1-\theta} \tag{1255}$$

The steady state solution also yields this: use $Y_{t+1} = Y_t = Y$ and $X_{t+1} = X_t = X$ then $Y_{t+1} = \delta + \theta Y_t + \phi_0 X_{t+1} + \phi_1 X_t + u_{t+1}$ becomes

$Y = \delta + \theta Y + \phi_0 X + \phi_1 X ==>$

$$Y = \frac{\delta}{1-\theta} + \frac{\phi_0 + \phi_1}{1-\theta} X \tag{1256}$$

This model also has error correction representation

$$\Delta Y_t = \delta - (1-\theta)\left(Y_{t-1} - X_{t-1}\right) + \phi_0 \Delta X_t + \gamma X_{t-1} + u_t \tag{1257}$$

where $\gamma = \phi_0 + \phi_1 - \theta - 1$

More generic ARDL(p,q) model:

$$\theta(L) Y_t = \delta + \phi(L) X_t + u_{t+2} \tag{1258}$$

$\theta(L) = 1 - \theta_1 L - \theta_2 L - ..... - \theta_p L^p$ and $\phi(L) = 1 + \phi_1 L + \phi_2 L + ... + \phi_p L^q$

Long run and short run multipliers

More generic ARDL(p,q) model:

$$\theta(L) Y_t = \delta + \phi(L) X_t + u_{t+2} \tag{1259}$$

$\theta(L) = 1 - \theta_1 L - \theta_2 L - ..... - \theta_p L^p$

$\phi(L) = 1 + \phi_1 L + \phi_2 L + ... + \phi_p L^q$

$$Y_t = \theta^{-1}(1)\,\delta + \theta^{-1}(L)\,\phi(L)\,X_t + \theta^{-1}(L)\,u_{t+2} \tag{1260}$$

$$\theta^{-1}(L)\,\phi(L) = \frac{1 + \phi_1 + \phi_2 + ... + \phi_p}{1 - \theta_1 - \theta_2 - ..... - \theta_p} \tag{1261}$$

$\theta^{-1}(L)\,\phi(L)$ measures the impact of X_t on Y_t , given that $\theta_1 + \theta_2 + ... + \theta_p < 1$, this series is invertible.

6.1.10 Vector Error Correction Model (VECM)

This now should help to understand the Johansen Procedure in cointegration. As mentioned before.

Start with X_t be vector of $N \times 1$ dimension each integrated of order 1 $I(1)$. The VAR is given by

$X_t = \Pi_1 X_{t-1} + \Pi_2 X_{t-2} + ... + \Pi_k X_{t-k} + \varepsilon_t$

Π_i is $N \times N$ matrix of coefficients This implies

$\begin{pmatrix} x_{1,t} \\ . \\ . \\ x_{n,t} \end{pmatrix} = \begin{pmatrix} \Pi_{111} & \Pi_{112} & . & \Pi_{11N} \\ \Pi_{121} & . & . & \Pi_{121} \\ . & . & . & . \\ \Pi_{1N1} & \Pi_{N21} & . & \Pi_{1NN} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ . \\ . \\ x_{n,t-1} \end{pmatrix} + .. + \begin{pmatrix} \Pi_{11k} & \Pi_{12k} & . & \Pi_{1Nk} \\ \Pi_{21k} & . & . & \Pi_{21k} \\ . & . & . & . \\ \Pi_{N1k} & \Pi_{N2k} & . & \Pi_{NNk} \end{pmatrix} \begin{pmatrix} x_{1,t-k} \\ . \\ . \\ x_{n,t-k} \end{pmatrix}$
$+ \begin{pmatrix} e_{1,t} \\ . \\ . \\ e_{n,t} \end{pmatrix}$

ECM representation of this VAR, popularly called VECM is given by

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1262}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_1 + ... + \Pi_i$ for $i = 1..k$

Γ_k gives the long run solution.

ECM representation of this VAR, popularly called VECM is given by

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1263}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_1 + ... + \Pi_i$ for $i = 1..k$

Γ_k gives the long run solution.

Instructions for PCGive. Load data in *.csv format. Then use the Econometrics package/multiple equation modelling/ select unrestricted system/ choose lags/determine the rank of cointegrating vector/ examine the Johansen test to determine the cointegrating vector./them do the cointegrated VAR/ analyse long run equilibrium relations and short run adjustment coefficients. Eviews has a good routine to compute this too.

6.1.11 Exercise: Empirical Analysis in Trade and Exchange Rate Model with Structural VAR

- Structural VAR is popular for this type of study as it allows to put restrictions based on the theoretical predictions in the above model.
- We limit our analysis to five variables that include relative wage between China and the US (w_{cu}), interest rate differential between China and the US (r_{cu}), Chinese real effective exchange rate (e), US relative GDP between China and the US (ry_{cu}) and the current account balance (CA_u) .
- The raw time series of these data are presented in Figure 1. When the Chinese economy has been growing rapidly, the exchange rate being fixed leads us to use China's real exchange rate rather than the nominal exchange rate. By doing so, we are also capturing the relative price effect.
- China's unit labour cost (ULC) is measured as total wage bill over real output (nominal output divided by CPI (1985=100)). Then relative wage is calculated by dividing ULC-China over ULC-US.

- Relative GDP on the other hand has been defined as Chinese GDP in dollar terms over US GDP. We calculate interest rate differential as the difference between Chinese average inter-bank rate and US 3-month Tbill rate. Current account balance for the US is used as the percentage of US nominal GDP.
- With these five variables, we formulate a first -order structural VAR of the following form:

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t-1} \\ r_{cu,t-1} \\ e_{c,t-1} \\ ry_{cu,t-1} \\ CA_{u,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{ryt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1264}$$

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR
where matrix notations can be employed for more exact representation.

$$X_t = \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix}; \ X_{t-1} = \begin{bmatrix} w_{cu,t-1} \\ r_{cu,t-1} \\ e_{c,t-1} \\ ry_{cu,t-1} \\ CA_{u,t-1} \end{bmatrix}; \ \varepsilon_t = \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{ryt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1265}$$

or compactly the path of X_{it} is affected by both contemporaneous and lagged effects of X_{jt} as measured by Γ_0 and Γ_1 and its own past values .

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR
Consider

$$X_t = B^{-1}\Gamma_0 + B^{-1}\Gamma_1 X_{t-1} + B^{-1}\varepsilon_t \tag{1266}$$

$$B^{-1} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix}; \ \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix}$$

The reduced form of this VAR system is then given by:

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1267}$$

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR
where $A_0 = B^{-1}\Gamma_0$, $A_1 = B^{-1}\Gamma_1$, $e_t = B^{-1}\varepsilon_t$

Reduced form is estimated with the available data; then structural shocks are retrieved using $e_t = B^{-1}\varepsilon_t$ This requires estimation of the variance covariance matrix of the error term

$$\sum = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} & \sigma_{35} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} & \sigma_{45} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \end{bmatrix} \tag{1268}$$

where $\sigma_{ij} = \frac{1}{T}\sum\limits_{t=1}^{T} e_{ij}e_{ij}^{'}$

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR

VAR is a-theoretic. In order to understand the long-run dynamics, we perform impulse response shock analysis, as the results from impulse responses are more informative than the estimated VAR regression coefficients (see Stock and Watson, 2001). It is customary to impose restrictions on coefficients based on prior economic theory. These restrictions can be on parameters, variance covariance matrices or symmetry.

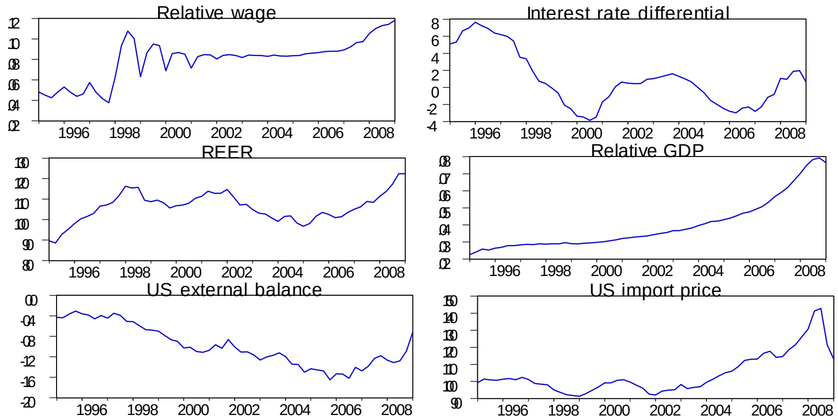
$$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1269}$$

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR

- Quarterly observations from 1995-Q1 to 2009-Q1 are used to estimate the model with two optimal lags.
- All the data have been gathered from Datastream and the variables are plotted in Figure 1.
- Since there is evidence of a structural break around 1994Q1 in China (see for example Baak (2008)), our sample in this paper starts from 1995Q1.
- Furthermore there is unavailability of quarterly data for the variables involved in this paper prior to 1995Q1.

Empirical Analysis in Trade and Exchange Rate Model with Structural VAR

Figure 1: Plot of time series used in the VAR



Impulse Response Analysis

The VAR is formulated with the following ordering: relative wage, interest rate differential, Chinese REER, relative GDP, and US current account balance. Shocks are extracted by applying a recursive identification structure with the above ordering. All the estimations have been carried out using RATS econometric software.

$$\begin{bmatrix} w_{c,t} \\ r_{c,t} \\ e_{c,t} \\ ry_{u,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \overline{w}_{c,t} \\ \overline{r}_{c,t} \\ \overline{e}_{c,t} \\ \overline{ry}_{u,t} \\ \overline{CA}_{u,t} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \\ e_{3t-i} \\ e_{4t-i} \\ e_{5t-i} \end{bmatrix} \tag{1270}$$

Impulse Response Analysis

Errors of the reduced form equations are related to the structural parameters as:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \\ e_{5t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1271}$$

Impulse Response Analysis

Introducing more simplifying assumptions:

$$\begin{bmatrix} w_{c,t} \\ r_{c,t} \\ e_{c,t} \\ pm_{u,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \overline{w}_{c,t} \\ \overline{r}_{c,t} \\ \overline{e}_{c,t} \\ \overline{pm}_{u,t} \\ \overline{CA}_{u,t} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11} & \phi_{12} & \phi_{13} & \phi_{14} & \phi_{15} \\ \phi_{21} & \phi_{22} & \phi_{23} & \phi_{24} & \phi_{25} \\ \phi_{31} & \phi_{32} & \phi_{33} & \phi_{34} & \phi_{35} \\ \phi_{41} & \phi_{42} & \phi_{43} & \phi_{44} & \phi_{45} \\ \phi_{51} & \phi_{52} & \phi_{53} & \phi_{54} & \phi_{55} \end{bmatrix} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1272}$$

Impulse Response Analysis

where $\phi_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1}$

$\phi_{i,j}(n)$ are impulse response coefficients for equation n. More compactly this can be represented as:

$$x_t = \mu + \sum_{i=0}^{\infty} \phi_i(i)\epsilon_{t-i} \tag{1273}$$

Exercise estimate the VAR, compute impulse responses and illustrate variance decomposition.

6.1.12 Restrictions and Variance Decomposition

To further validate this result, a 6-variable VAR has been formulated by adding US import price as another variable in the VAR, following an over-identified SVAR strategy (Sims-Zha) and impose the restrictions in the matrix below:

$$\begin{bmatrix} 1 & a_{12} & 0 & 0 & a_{15} & 0 \\ 0 & 1 & a_{23} & 0 & a_{25} & 0 \\ a_{31} & a_{32} & 1 & a_{34} & a_{35} & a_{36} \\ a_{41} & 0 & a_{43} & 1 & 0 & a_{46} \\ a_{51} & a_{52} & 0 & 0 & 1 & a_{56} \\ 0 & 0 & a_{63} & a_{64} & 0 & a_{66} \end{bmatrix}$$

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1274}$$

from successive iteration this reduces to

$$E_t X_{t+n} = \left(I + A_1 + A_1^2 + A_1^3 + ... + A_1^{n-1}\right) A_0 + A_1^n X_t + e_t \tag{1275}$$

Forecast error is given by

$$\left(e_{t+n} + A_1 e_{t+n-1} + A_1^2 e_{t+n-2} + ... + A_1^{n-1} e_{t+1}\right) A_0 + A_1^n X_t \tag{1276}$$

$$X_{t+n} - E_t X_{t+n} = \sum_{i=0}^{n-1} \phi_i(i)\epsilon_{t+n-i} \tag{1277}$$

Taking only one equation

$w_{t+n} - E_t w_{t+n} = \phi_{11}(0)\epsilon_{wt+n} + \phi_{11}(1)\epsilon_{wt+n-1} + ... + \phi_{11}(n-1)\epsilon_{wt+1} + \phi_{12}(0)\epsilon_{rt+n} + \phi_{12}(1)\epsilon_{rt+n-1} + ... + \phi_{12}(n-1)\epsilon_{rt+1}$

$+\phi_{12}(0)\epsilon_{et+n} + \phi_{12}(1)\epsilon_{et+n-1} + ... + \phi_{12}(n-1)\epsilon_{et+1}$

$+\phi_{12}(0)\epsilon_{pmt+n} + \phi_{12}(1)\epsilon_{pmt+n-1} + ... + \phi_{12}(n-1)\epsilon_{pmt+1}$

$+\phi_{12}(0)\epsilon_{cat+n} + \phi_{12}(1)\epsilon_{cat+n-1} + ... + \phi_{12}(n-1)\epsilon_{cat+1}$

Variance of n-step ahead forecast error is

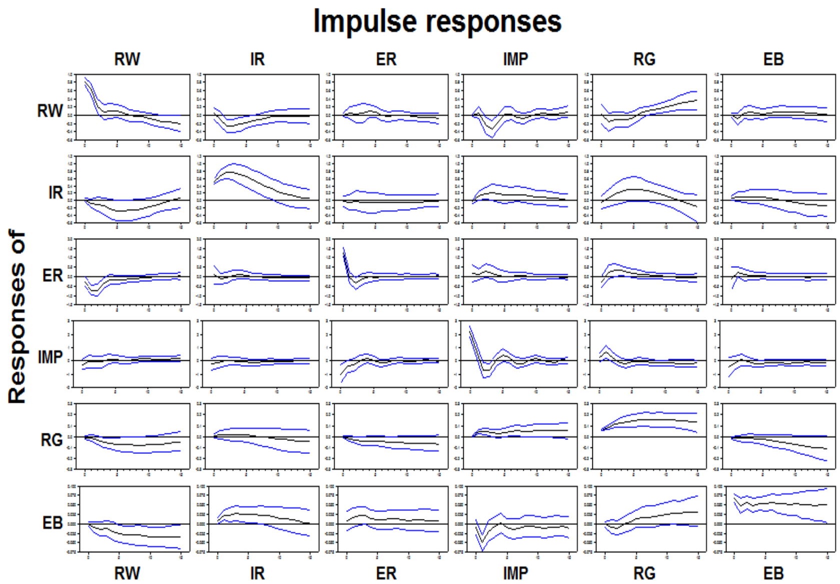
$$\begin{aligned}\sigma(n)_w^2 \quad = \quad & \sigma_w^2\left[\phi_{11}(0)+\phi_{11}(1)+...+\phi_{11}(n-1)\right]+\\ & \sigma_r^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]+\\ & \sigma_e^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]+\\ & \sigma_{pm}^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]+\\ & \sigma_{CA}^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right] \qquad (1278)\end{aligned}$$

Variance decomposition in terms of variances of shocks ϵ_{wt} , ϵ_{rt}, ϵ_{et}, ϵ_{pmt} and ϵ_{cat}.

$$\begin{aligned}\sigma(n)_w^2 \quad = \quad & \frac{\sigma_w^2\left[\phi_{11}(0)+\phi_{11}(1)+...+\phi_{11}(n-1)\right]}{\sigma(n)_w^2}+\\ & \frac{\sigma_r^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]}{\sigma(n)_w^2}+\\ & \frac{\sigma_e^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]}{\sigma(n)_w^2}+\\ & \frac{\sigma_{pm}^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]}{\sigma(n)_w^2}+\\ & \frac{\sigma_{CA}^2\left[\phi_{12}(0)+\phi_{12}(1)+...+\phi_{12}(n-1)\right]}{\sigma(n)_w^2} \qquad (1279)\end{aligned}$$

Thus the variance decomposition is finding the proportion of variance explained by a variable's own shock(ϵ_{wt}) versus the variance explained by shock to the other variables ϵ_{rt}, ϵ_{et}, $\epsilon_{pmt}$and ϵ_{cat}. The variance decomposition for the empirical model is presented in Table 3. In the variance decomposition analysis, nearly 75% of the variation in US current account balance is explained by its own shocks, and relative GDP explains 29% of the variation in relative wage. As nearly 25% of the variation in US current account balance is explained by interest rate differential (11%), REER (5%), relative GDP (6%) and China's wage cost (3%), this could suggest that China's exchange rate appreciation might not solve the enlarging US current account deficits. However from the long-run and short-run parameter estimates, higher relative GDP of China does have a significant effect on lowering current account balance, and from variance decomposition results, 12% of the variation in relative GDP is on the back of China's relatively lower wage cost. Figure 2 shows that following a relative wage shock, relative GDP declines with either loss of income for the low-wage country or the rise in income for the high-wage country.

Figure 7: IRFs from over-identified SVAR



Structural Coefficients

$$\begin{bmatrix} w_{cu,t} \\ r_{cu,t} \\ e_{c,t} \\ ry_{cu,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1280}$$

Structural Coefficients

$$\begin{bmatrix} a_{10} \\ a_{20} \\ a_{30} \\ a_{40} \\ a_{50} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \\ b_{40} \\ b_{50} \end{bmatrix} ; \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \\ e_{5t} \end{bmatrix} \tag{1281}$$

$$= \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1282}$$

Structural Coefficients

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \gamma_{25} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \gamma_{45} \\ \gamma_{51} & \gamma_{52} & \gamma_{53} & \gamma_{54} & \gamma_{55} \end{bmatrix} \tag{1283}$$

Structural Coefficients

$$\begin{bmatrix} w_{c,t} \\ r_{c,t} \\ e_{c,t} \\ pm_{u,t} \\ CA_{u,t} \end{bmatrix} = \begin{bmatrix} \overline{w}_{c,t} \\ \overline{r}_{c,t} \\ \overline{e}_{c,t} \\ \overline{pm}_{u,t} \\ \overline{CA}_{u,t} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix}^{-1} \tag{1284}$$

$$\begin{bmatrix} \varepsilon_{wt} \\ \varepsilon_{rt} \\ \varepsilon_{et} \\ \varepsilon_{pmt} \\ \varepsilon_{cat} \end{bmatrix} \tag{1285}$$

Eviews 8 allows to compute the Baysian VAR.

References

[1] Athanasopoulos G and F. Vahid (2008) VARMA versus VAR for Macroeconomic Forecasting Journal of Business & Economic Statistics, 26, 2, 237-252

[2] Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. Journal of Business and Economic Statistics 23, 49–60.

[3] Benati L and P. Surico (2009) VAR Analysis and the Great Moderation, American Economic Review, 99, 4, 1636-1652

[4] Bernanke Ben (1986) Alternative explantions of money income correlations, Canegie Rochester Conference Series on Public Policy, 25, 49-100. http://econ.la.psu.edu/~hbierens/EasyRegTours/VAR.HTM

[5] Bhattarai K and S. Mallick (2013) Impact of China's currency valuation and labour cost on the US in a trade and exchange rate model, North American Journal of Economics and Finance, 25:40-59.

[6] Blanchard O and D Quah (1989) The dynamic effects of aggregate demand and supply disturbances, American Economic Review, 79, 655-673.

[7] Canova, F and G. De Nicoló (2002), "Monetary Disturbances Matter for Business Fluctuations in the G–7." Journal of Monetary Economics, 49(6): 1131–59.

[8] Chahrour, R, S. Schmitt-Grohé, and M. Uribe (2012) A Model-Based Evaluation of the Debate on the Size of the Tax Multiplier, American Economic Journal: Economic Policy, 4(2): 28-45

[9] Davidson, R. (2012), Statistical inference in the presence of heavy tails, Econometrics Journal, 15: C31–C53

[10] Davidson R and MacKinnon J. G. (2004) Econometric Theory and Methods, Oxford.

[11] Dees S., F. Di Mauro, M. H. Pesaran and L. V. Smith (2007) Exploring the International Linkages of the Euro Area: A Global Var Analysis, Journal of Applied Econometrics, 22, 1, 1-38

[12] Dowd K and D. Blake (2006) After VaR: The Theory, Estimation, and Insurance Applications of Quantile-Based Risk Measures, Journal of Risk and Insurance, 73, 2,193-229

[13] Dungey, M. and R Fry (2009) “The Identification of Fiscal and Monetary Policy in a Structural VAR, Economic Modelling, 26(6):1147–60.

[14] Enders W and S Hurn (2006) Identifying aggregate demand and supply shocks in a small open economy Oxford Economic Papers, 1-19

[15] Enders W (2010) Applied Econometric Time Series, John Wiley

[16] Fair R C (1984) Specification, estimation, and analysis of macroeconometric models, Harvard.

[17] Fanelli L (2007) Present Value Relations, Granger Noncausality, and VAR Stability Econometric Theory, 23, 6,1254-1260

[18] Favero, C., and F. Giavazzi. 2012. "Measuring Tax Multipliers: The Narrative Method in Fiscal VARs." American Economic Journal: Economic Policy, 4(2): 69-94.

[19] Fernández-Villaverde, J, J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson. 2007. "ABCs (and Ds) of Understanding VARs." American Economic Review, 97(3): 1021-1026

[20] Fry, Renée, and A. Pagan (2011) Sign Restrictions in Structural Vector Autoregressions: A Critical Review, Journal of Economic Literature, 49(4): 938-60

[21] Hamilton JD (1994) Time Series Analysis, Princeton.

[22] Harris R and R Sollis (2003) Applied Time Series Modelling and Forecasting, John Wiley and Sons.

[23] Hendrry David (1995) Dynamic Econometrics, Oxford University Press.

[24] Hendry D.F and J. A. Doornik (1994) Modelling linear dynamic econometric Systems, Scottish Journal of Polifical Economy, 41. I, 1-33

[25] Holly S. and M Weale ed. (2000) Econometric Methods: Practice and Applications, Cambridge University Press.

[26] Garratt, A., Lee, K., Pesaran, M.H. and Shin, Y. (2003) ‘A structural cointegration VAR approach to macroeconometric modelling’, Economic Journal, Vol. 113, No. 487, pp.412–445

[27] Hendry D. F (1974) Stochastic Specification in an Aggregate Demand Model of the United Kingdom, Econometrica, 42, 3, 559-578

[28] Hendry D. F (2002) Forecast Failure, Expectations Formation and the Lucas Critique, Annals of Economics and Statistics,The Econometrics of Policy Evalution, 21-40

[29] Judge G G, R.C. Hill, W.E. Griffiths, H Lutkepohl and TC Lee (1988) Introduction to the Theory and Practice of Econometrics, John Wiley

[30] Kadiyala K. R. and S. Karlsson (1997) Numerical Methods for Estimation and Inference in Bayesian VAR-Models, Journal of Applied Econometrics, 12, 2, 99-132

[31] Kapetanios, G, A Pagan, and A. Scott (2007) Making a Match: Combining Theory and Evidence in Policy-Oriented Macroeconomic Modeling, Journal of Econometrics, 136(2): 565–94.

[32] Kocherlakota N. R. and K. M Yi (1996) A simple time series tests of endogenous vs. exogenous growth models : An application to the United States, Review of Economic Studies, 78:1:126-134.

[33] Luintel, K.B. and Khan, M., 1999. A quantitative reassessment of the finance–growth nexus: evidence from a multivariate VAR. Journal of development economics, 60(2), pp.381-405.

[34] Lucas R E and T. J Sargent (1981) Rational Expectations and Econometric Practice, University of Minnesota Press.

[35] Mills TC (2000) Econometric Modelling of Financial Time Series, 2nd ed. Cambridge

[36] Sims C. A. (1980) Macroeconomics and Reality, Econometrica, 48:1: 1-48.

[37] Swensen A R (2006) Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models, Econometrica, 74, 6 ,1699-1714

[38] Wallis K.F. (1989) Macroeconomic Forecasting: A Survey, Economic Journal, 99, March, 28-61.

[39] http://www.feweb.vu.nl/econometriclinks/

[40] Mertens K and M. O. Ravn (2010) Measuring The Impact Of Fiscal Policy In The Face Of Anticipation: A Structural Var Approach The Economic Journal, 120, 544, 393-413

[41] Mountford, A (2005) Leaning into the Wind:A Structural VAR Investigation of UK Monetary Policy, Oxford Bulletin of Economics and Statistics, 67(5): 597–621.

[42] Pagan A. and M. Wickens (1989) A Survey of Some Recent Econometric Methods, Economic Journal, 99:962-1025.

[43] Patterson K (2010) A Primer for Unit Root Testing, Palgrave macmillan.

[44] Patterson K (2000) An Introduction Applied Econometrics: A Time Series Approach,MacMillan.

[45] Phillips PCB and T. Magdalinos (2013) Inconsistent Var Regression With Common Explosive Roots, Econometric Theory, 29, 4, 808-837

[46] Mills TC (2000) Econometric Modelling of Financial Time Series, 2nd ed. Cambridge

[47] Rafiq, M.S. and S.K. Mallick (2008) The effect of monetary policy on output in EMU3: A sign restriction approach, *Journal of Macroeconomics*, 30 (4): 1756-1791.

[48] Sims C. A. (1980) Macroeconomics and Reality, Econometrica, 48:1: 1-48.

[49] Sims, C.A. and T. Zha (1999) Error Bands for Impulse Responses, *Econometrica*, 67 (5): 1113-1155

[50] Stock J. H. and M. W. Watson (2005) Understanding changes in international business cycle dynamics,Journal of European Economic Association, 3:5:968-1006.

[51] Stock, J.H., and M.W. Watson (2001). Vector Autoregressions, *Journal of Economic Perspectives*, 15 (4): 101-115.

[52] Vargas-Silva, Carlos. 2008. "Monetary Policy and the US Housing Market: A VAR Analysis Imposing Sign Restrictions." Journal of Macroeconomics, 30(3):977–90.

[53] Wallis K.F. (1989) Macroeconomic Forecasting: A Survey, Economic Journal, 99, March, pp 28-61.

[54] http://www.feweb.vu.nl/econometriclinks/

7 L6: Cross Section Analyis

Cross section analysis is applied widely in economics and related fields. Labour economists use for analysing earning and labour supply functions; participation rates, wage rates. They attempt to measure not only the impact of economic factors on participation, but also to show how personal beliefs and political alliance are related to willingness to work, personal perception of the world and the local environment and set of educational and psychological profiles of an individual can influence the decision on whether to enter or remain outside the labour force.

- And indicative literature in cross section analysis:
- Tobin (1958), Theil(1969),Hausman (1978), Heckman (1979), Ashenfelter (1986), Lancaster and Chesher (1983), Maddala (1983),
- McFadden (1974), Staigler, Stock, (1997), Ruud (2000) Verbeek (2004), Greene (2008),Wooldridge (2002), Güell and Hu (2006), Phillips and Sul (2007), Chesher and Rosen (2013)
- Blundell and Smith (1989), Chesher (1984),Smith (1997), Greene (2000), Verbeek (2012),Imbens and Lancaster (1994), Keifer (1988), Hey and Orme (1994) ,Staigler and Stock (1997) ,Blundell(2013)
- White (1980), Cowing and Holtmann (1983), Baltagi (1984), Stoker(1986), Nuamah(1992), Moffitt (1993), Coles and Smith (1996), Amacher and Hellerstein (1999), Perali and Chavas (2000), Gomes, Kogan, and Zhang (2003), Alvarez and Arellano (2003), Higson, Holly, Kattuman and Platis (2004), Andrews (2005), Madsen (2005), Güel and Hu (2006), Phillips and Sul (2007), Pesaran, Ullah and Yamagata (2008), Kapetanios (2008), Chudik, Pesaran and Tosetti (2011), Bai and Ng (2010),Chesher (2010) Chesher and Rosen (2013)

Cross section techniques are very effectively applied in modelling probabilities of occurence or non-occurence of certain event depending upon a set of observed and unobserved determining factors. Microecomists use cross section analysis to study behaviour of consumers and producers in the market, to determine demand or supply, revenue, sales or profit; survival or hazard rates in the business. Macroeconomists use cross section anlysis for compresensive and comparative study on growth rates, employments or borrowing, lending or credit rationing problems. Probability strokes or effectivness of certain experiments in health; treatment and control group in experimental economics (Hey and Orme (1994)). These techniques are essential in public policy analysis (Blundell , Pistaferri, and Preston (2008); Atkinson and Brandolini (2010)) or in sports economics (Dobson and Goddard (2011). While cross section techniques most ofter are applied to very rich set of cross section data, often millions of observations of census or survey or panel dataset, there are specifice issues including the truncation, corretion for sample selection bias, instrumental variables to account for unobserved heterogeniety. This section will provide initial models of cross section analysis, more sofisticated models can be found the Handbook of Econometrics Volumes 1 to 6B (eds. Heckmand and Leamer, McFadden; Zvi Griliches, Robert F. Engle, Michael D. Intriligator) or in the Handbook of Labour Economics.

7.0.13 Dummy Dependent Variable Regression Model

- Alternative names: dichotomous dependent variables, discrete dependent random variable, binary variable, either or choice variables

$$Y_i = \int \begin{array}{l} Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad \text{if the event occurs} \\ 0 = \text{otherwise} \end{array} \tag{1286}$$

  Examples

- the labour force participation (1 if a person participates in the labour force, 0 otherwise)
- yes or no vote in particular issue ; to marry or not to marry; to study further or to start a job
- to buy or not to buy a particular stock
- choice of transportation mode to work (1 if a person drives to work, 0 otherwise)
- Union membership (1 if one is a member of the union, 0 otherwise)
- Owning a house (1 if one owns 0 otherwise)
- Multinomial choices: work as a teacher, or as a clerk, or as a self employed or professional or as a factory worker
- Multinomial ordered choices: strongly agree, agree, neutral, disagree

Best way to learn crosss section analysis and probability models is to follow series of lab-exercises developed by Professor William Greene of the New York University (http://people.stern.nyu.edu/wgreene/). While the stories and analytical derivations are in Greene's text book Econometric Analysis (7th edition), the details of workshop based course in Hull between July 1-2, 2013 can be found in his web page http://people.stern.nyu.edu/wgreene/Hull2013.htm. These contains both derivations and computations of microeconometric analysis of random utility models, multiple choice and nested logit models, random parameter multinomial logit models and duration models.

7.0.14 Linear Probability Model

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{1287}$$

where $Y_i = 1$ if person owns a house, 0 otherwise; X_i is family income.
$E\left[(Y_i = 1)/X_i\right]$ probability that the event y will occur given x

$$E\left[(Y_i = 1)/X_i\right] = 0 \times [1 - P_i] + 1 \times P_i = P_i \tag{1288}$$

$$0 \leqslant E\left[(Y_i = 1)/X_i\right] = P_i = \beta_1 + \beta_2 X_i \leqslant 1 \tag{1289}$$

- Problem: Errors are heteroskedastic.

$$\varepsilon_i = 1 - \beta_1 - \beta_2 X_i \quad with \quad (1 - P_i) \tag{1290}$$

$$\varepsilon_i = -\beta_1 - \beta_2 X_i \quad with \quad P_i \tag{1291}$$

Variance of error in a linear probability model

$$var\left(\varepsilon_i\right) = \left(1 - \beta_1 - \beta_2 X_i\right)^2 \; (1 - P_i) + \left(-\beta_1 - \beta_2 X_i\;\right)^2 P_i \tag{1292}$$

$$\sigma^2 = \left(1 - \beta_1 - \beta_2 X_i\right)^2 \; \left(-\beta_1 - \beta_2 X_i\right) + \left(-\beta_1 - \beta_2 X_i\;\right)^2 \left(1 - \beta_1 - \beta_2 X_i\right) \tag{1293}$$

$$\sigma^2 = \left(1 - \beta_1 - \beta_2 X_i\right)\left(\beta_1 + \beta_2 X_i\right) \; = (1 - P_i) \; P_i \tag{1294}$$

Variance depends on X.
Limitations of a linear probability model
It is possible to transform this model to make it homeskedastic by dividing the original variables by

$$\sqrt{\left(1 - \beta_1 - \beta_2 X_i\right)\left(\beta_1 + \beta_2 X_i\right)} \; = \sqrt{(1 - P_i) \; P_i} = \sqrt{W_i} \tag{1295}$$

$$\frac{Y_i}{\sqrt{W_i}} = \frac{\beta_1}{\sqrt{W_i}} + \beta_2 \frac{X_i}{\sqrt{W_i}} + \frac{\varepsilon_i}{\sqrt{W_i}} \tag{1296}$$

- It does not guarantee that the probability lies inside (0,1) bands
- Probability in non-linear phenomenon: at very low level of income a family does not own a house; at very high level of income every one owns a house ; marginal effect of income is very negligible. The linear probability model does not explain this fact well.

7.0.15 Probit Model

-

$$\begin{aligned}
\Pr\left(Y_i = 1\right) &= \Pr\left(Z_i^* \leq Z_i\right) = F\left(Z_i\right) = \frac{1}{\sqrt{2\pi}} \int \begin{matrix} Z_i \\ -\infty \end{matrix} \; e^{-\frac{t^e}{2}} dt \\
&= \frac{1}{\sqrt{2\pi}} \int \begin{matrix} \beta_1 + \beta_2 X_i + \varepsilon_i \\ -\infty \end{matrix} \; e^{-\frac{t^e}{2}} dt
\end{aligned} \tag{1297}$$

- Here t is standardised normal variable, $t \sim N\left(0, 1\right)$

  probability depends upon unobserved utility index Z_i which depends upon observable variables such as income. There is a thresh-hold of this index when after which family starts owning a house, $Z_i \geqslant Z_i^*$
- The Unknown parameters of this model α, β and σ are estimated by the maximum likelihood estimator.

Probit Model

- Likelihood functions for modelling probability

$$L = Pr\left(Y_1\right) Pr\left(Y_2\right) ...Pr\left(Y_N\right) \tag{1298}$$

$$L = \prod_{i=1}^{n1} P_i \prod_{i=n_{1+1}}^{N} \left(1 - P_i\right) \tag{1299}$$

$$L = \prod_{i=1}^{n1} P_i^{y_i} \prod_{i=n_{1+1}}^{N} \left(1 - P_i\right)^{(1-y_i)} \tag{1300}$$

Log likelihood functions for modelling probability

$$\log L = \sum_{i=1}^{n1} \log P_i \sum_{i=n_{1+1}}^{N} \log\left(1 - P_i\right) \tag{1301}$$

Probit Model

- First order conditions for ML estimates of and

$$\frac{\log L}{\partial \alpha} = \sum_{i=1}^{n1} \frac{\partial P_i / \partial \alpha}{P_i} - \sum_{i=n_{1+1}}^{N} \frac{\partial P_i / \partial \alpha}{\left(1 - P_i\right)} = 0 \tag{1302}$$

$$\frac{\log L}{\partial \beta} = \sum_{i=1}^{n1} \frac{\partial P_i / \partial \beta}{P_i} - \sum_{i=n_{1+1}}^{N} \frac{\partial P_i / \partial \beta}{\left(1 - P_i\right)} = 0 \tag{1303}$$

ML takes the OLS estimates as starting value for computations of optimal α, β and σ.

$R^2 = 1 - \frac{\log L_{LR}}{\log L_{UR}}$ gives the indication of goodness of fit of ML estimates. This is the proportion of correct predictions.

MLE Procedure

- $$L\left(\alpha, \beta, \sigma\right) = L\left(y_1, y_2, ...y_N\right) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{\left(y_i - \alpha - \beta X_i\right)^2}{\sigma^2}\right] \tag{1304}$$

Take log of this function to get a log-liklihood function.

$$\begin{aligned} \log L\left(\alpha, \beta, \sigma\right) &= \sum_{i=1}^{N} - \frac{1}{2} \log\left(2\pi\sigma^2\right) - \frac{1}{2} \sum_{i=1}^{N} \frac{\left(y_i - \alpha - \beta X_i\right)^2}{\sigma^2} \\ &= c - \frac{N}{2} \log\left(\sigma^2\right) - \frac{Q}{2\sigma^2} \end{aligned} \tag{1305}$$

where $c = \frac{N}{2} \log\left(2\pi\right)$ and $Q = \sum_{i=1}^{N} \left(y_i - \alpha - \beta X_i\right)^2$

Maximising this likelihood w.r.t. α, β and σ. is equivalent to minimizing Q , which is the negative term in the likelihood function. Therefore the estimators of α, β and σ under the ML method are the same as in the OLS method.

7.0.16 Logit Model

- variable Y_i which takes value 1 ($Y_i = 1$) if a student gets a first class mark, value 0 ($Y_i = 0$)otherwise.
- Probability of getting a first class mark in an exam is a function of student effort index denoted by .Z_i ; where $P_i = \frac{1}{1+e^{-Z_i}}$

  $Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ An example of a logit model: what determines that a student gets the first class degree?

$$Z_i = \beta_1 + \beta_2 H_i + \beta_3 E_i + \beta_4 A_i + \beta_2)_i + \varepsilon_i \tag{1306}$$

  H = hours of study, E= exercises, A = attendance in lectures and classes; P = papers written for assignment.

- Ratio of odds: $\frac{P_i}{1-P_i} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i}$; taking log of the odds $ln\left(\frac{P_i}{1-P_i}\right) = Z_i$

Features of a logit Model

- - probability goes from 0 to 1 as the index variable goes from -∞ to +∞. Probability lies between 0 and 1.
  - Log of the odds is linear in x, characteristic variables but probabilities themselves are not linear but non linear function of the parameters. Probabilities are estimated using the maximum likelihood method.
  - Any explanatory variable that determines the value of Z_i , measures how the log of odds of an event (i.e. owning a house) changes as a result of change in explanatory variable such as income.
  - We can calculate P_i for given estimates of $\beta_1$and $\beta_2$or all other .β_i
  - Limiting case when P_i =1; $ln\left(\frac{P_i}{1-P_i}\right)$ or when P_i =0 ; $ln\left(\frac{0}{1-0}\right)$OLS cannot be applied in such case but the maximum likelihood method may be used to estimate the parameters.

Features of a logit Model

$$P_i = \frac{1}{1+e^{-Z_i}} \tag{1307}$$

$$Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{1308}$$

$$1 - P_i = 1 - \frac{1}{1+e^{-Z_i}} = \frac{1+e^{-Z_i}-1}{1+e^{-Z_i}} = \frac{e^{-Z_i}}{1+e^{-Z_i}} = \frac{1}{1+e^{Z_i}} \tag{1309}$$

e = exp = 2.718

$$\frac{P_i}{1-P_i} = \frac{\frac{1}{1+e^{-Z_i}}}{\frac{1}{1+e^{Z_i}}} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = \frac{e^{Z_i}\left(1+e^{Z_i}\right)}{e^{Z_i}\left(1+e^{-Z_i}\right)} = e^{Z_i} \tag{1310}$$

Taking log of the odds

$$ln\left(\frac{P_i}{1-P_i}\right) = Z_i \tag{1311}$$

Multinomial Choice Model

- choice of different brands of a particular goods such as cereals
- different subjects by students in the business school such as economics, marketing, finance, business, accounting, management or
- in the university such as science, engineering, medicine, mathematics, philosophy, politics or history, arts or like that.
- ordered probit or ordered logit for choice of bonds such as AAA BBB; orders are used to rank the outcome.
- survey questions with ranking.

Summary of Probability Models
The effect of observed variables on probability

- –

$$\frac{\partial P_i}{\partial x_{i,j}} = \begin{cases} \beta_j \\ \beta_j P_j\left(1-P_j\right) \\ \beta_j \phi\left(Z_i\right) \end{cases} \tag{1312}$$

  – where $Z_i = \beta_0 + \sum_{i=1}^{k} \beta_i X_{i,j}$ and ϕ is the standard normal density function.

7.0.17 Multinomial Choice Models

Let $P_{i,j}$ denote the probability of choosing alternative j by individual i,.

There are J alternatives and N individuals. According to McFadden (1974)

$$\sum_{i=1}^{n1} y_{i,j} = \sum_{i=1}^{n1} P_{i,j} = 1 \tag{1313}$$

Log likelihood

$$L = \prod_{i=1}^{n1} P_{i,1}^{y_{i,1}}.P_{i,2}^{y_{i,2}}.P_{i,3}^{y_{i,3}}....P_{i,J}^{y_{i,J}} \tag{1314}$$

Individuals make a choice to maximise utility; such utilities are specific to individual and unobserved factors .

$$u_{i,j} = \overline{u}_{i,j} + e_{i,j} = X'_{i,j}\ \beta + e_{i,j} \tag{1315}$$

Errors $e_{i,j}$ need to have Weibull distribution a multinomial logit.
Multinomial Logit with Random Utility Model
Probability that the choice 1 is made when J alternatives were available implies

$P_{i,j} = \Pr\left[u_{i,1} > u_{i,2} \text{ and } u_{i,1} > u_{i,3} \text{ ...and } u_{i,1} > u_{i,J}\right] =$
$P_{i,1} = \Pr\left[\begin{array}{c} e_{i,2} < u_{i,1} - u_{i,2} + e_{i,1} \text{ and } e_{i,3} < u_{i,1} - u_{i,3} + e_{i,1} \text{ .} \\ \text{..and } e_{i,J} < u_{i,1} - u_{i,J} + e_{i,1} \end{array}\right]$
$P_{i,1} = \Pr\left[\begin{array}{c} e_{i,2} - e_{i,1} < u_{i,1} - u_{i,2} \text{ and } e_{i,3} - e_{i,1} < u_{i,1} - u_{i,3} \text{ ..} \\ \text{.and } e_{i,J} - e_{i,1} < u_{i,1} - u_{i,J} \end{array}\right]$
Weibull errors: $\Pr\left[e_{i,j} < \varepsilon\right] = \exp\left(-\exp\left(-\varepsilon\right)\right)$

$$P_{i,1} = \frac{X'_{i,j}\ \beta}{\sum\limits_{j=1}^{J} X'_{i,j}\ \beta} \tag{1316}$$

Multinomial Logit and Independence of Irrelevant Alternatives

Choices are independent of irrelevant alternatives (IIA); only odds between two choices need to be compared when one is confronted with choosing one or another.

$$P_{i,1} = \frac{\exp\left(X'_{i,j}\ \beta\right)}{\sum\limits_{j=1}^{J} \exp\left(X'_{i,j}\ \beta\right)} \tag{1317}$$

$$\frac{P_{i,1}}{P_{i,2}} = \frac{\frac{\exp\left(X'_{i,2}\ \beta\right)}{\sum_{j=1}^{J} \exp\left(X'_{i,j}\ \beta\right)}}{\frac{\exp\left(X'_{i,1}\ \beta\right)}{\sum_{j=1}^{J} \exp\left(X'_{i,j}\ \beta\right)}} = \frac{\exp\left(X'_{i,2}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)} \tag{1318}$$

$$\frac{P_{i,k}}{P_{i,2}} = \frac{\exp\left(X'_{i,k}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)} = \exp\left(X'_{i,k}\ \beta - X'_{i,1}\ \beta\right) \quad \text{for k=2,....,J} \tag{1319}$$

Multinomial Logit and Independence of Irrelevant Alternatives

$$\frac{P_{i,k}}{P_{i,2}} = \frac{\exp\left(X'_{i,k}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)} = \exp\left(X'_{i,k}\ \beta - X'_{i,1}\ \beta\right) \quad \text{for k=2,....,J} \tag{1320}$$

with normalisation $\beta_1 = 0$;

$$P_{i,1} = \frac{\exp\left(X'_{i,j}\ \beta\right)}{\sum\limits_{j=1}^{J} \exp\left(X'_{i,j}\ \beta\right)} = \frac{1}{1 + \sum\limits_{j=2}^{J} \exp\left(X'_{i,j}\ \beta\right)} \quad \text{for j =2,.....,J} \tag{1321}$$

Parameters in a multinomial logit models are estimated by a maximum likelihood method (see McFadden (1974). Judge , Griffiths, Hill, Lutkepohl and. Lee(1990). Similarly one can define the nested logit models.

Nested Logit Models (see Greene: 23) If there are J alternatives and B branches a multiple logit evaluation of probability of choosing j in k subgroup is given by:

$$prob\,[twig, branch] = P_{ijb} = \frac{\exp\left(X'_{i,j\backslash b}\ \beta + z'_{i,b}\ \gamma\right)}{\sum\limits_{b=1}^{B}\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta + z'_{i,b}\ \gamma\right)} \tag{1322}$$

$$P_{ijb} = P_{ij\backslash b}P_b = \left(\frac{\exp\left(X'_{i,j\backslash b}\ \beta\right)}{\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta\right)}\right)\left(\frac{\exp\left(z'_{i,b}\ \gamma\right)}{\sum\limits_{l=1}^{L}\exp\left(z'_{i,b}\ \gamma\right)}\right) \tag{1323}$$

$$\left(\frac{\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta\right)\sum\limits_{l=1}^{L}\exp\left(z'_{i,b}\ \gamma\right)}{\sum\limits_{b=1}^{B}\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta + z'_{i,b}\ \gamma\right)}\right) \tag{1324}$$

Nested Logit Models (see Greene: 23)
This can be estimated using a two step Maximum likelihood procedure.

$$L = \sum_{j=1}^{N}\ln\left[pr\,(twig/branch)_i \times prob\,(branch)\right] \tag{1325}$$

Inclusive value of lth brach

$$IV_{ib} = \ln\left(\sum_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta\right)\right) \tag{1326}$$

$$P_{ij\backslash b}P_b = \left(\frac{\exp\left(X'_{i,j\backslash b}\ \beta\right)}{\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta\right)}\right); P_b = \frac{\exp\left(\tau_b\left(z'_{i,b}\ \gamma + IV_{ib}\right)\right)}{\sum\limits_{b=1}^{B}\exp\left(\tau_b z'_{i,b}\ \gamma + IV_{ib}\right)} \tag{1327}$$

Model can be estimated with limited information maximum likelihood (LIML)or full information maximum likelihood (FIML). (see Greene: 23)
Nested Logit Models (see Greene: 23)

$$P_{ij\backslash b}P_b = \left(\frac{\exp\left(X'_{i,j\backslash b}\ \beta\right)}{\sum\limits_{j=1}^{J_b}\exp\left(X'_{i,j\backslash b}\ \beta\right)}\right); P_b = \frac{\exp\left(\tau_b\left(z'_{i,b}\ \gamma + IV_{ib}\right)\right)}{\sum\limits_{b=1}^{B}\exp\left(\tau_b z'_{i,b}\ \gamma + IV_{ib}\right)} \tag{1328}$$

This can be estimated using a two step Maximum likelihood procedure.

It can be applied to model the choices of travel modes in which an individual chooses first whether to fly or take a ground transport.

Once a ground transport is chosen then similar choice is made for train, bus or car transport.

Model can be estimated with limited information maximum likelihood (LIML)or full information maximum likelihood (FIML). (see Greene: 23)

7.0.18 Count Data Models

Poisson random variable $P\left(Y=y\right)=\frac{e^{-\lambda}\lambda^{y}}{y!}$

where λ denotes the intensity of occurrence or the rate parameter and y = 1,2,,3,....denote the counts of events in a given time interval. Mean and variance are the same $E(Y)=\lambda$ and variance $var\left(Y\right)=\lambda$.

$\lambda_i=\exp\left(X_i'\ \beta\right)$

$$ln\left(\lambda_i\right)=X_i'\ \beta \tag{1329}$$

$$E\left(y_i/x_i\right)=var\left(y_i/x_i\right)=\lambda_i=e^{X_i'\ \beta} \tag{1330}$$

Likelihood function

$$\ln L=\left[-\lambda_i+y_i'X_i'\ \beta-\ln y_i!\right] \tag{1331}$$

Parameters is estimated by k number of first order conditions as is asymptotically normal with the sample covariance matrix

Gausss-Newton or Newton-Raphson or BHHH iterative algorithm is used to find unique parameters (see Green 25).

7.0.19 Ordered Probit Model

Ordered Probit Model (See Greene 831)

$$Y_i=\begin{cases}1 & if\ \ Y_i^*\leq 1\\ 2 & if\ \ 1<Y_i^*\leq 2\\ 3 & if\ \ 2<Y_i^*\leq 3\\ 4 & if\ \ 3<Y_i^*\leq 4\\ 5 & if\ \ 4<Y_i^*\leq 5_i\end{cases}$$

Ordered Probit Model

$$prob\,(y = 0|x) = \Phi\,(-x'\beta) \tag{1332}$$

$$prob\,(y = 1|x) = \Phi\,(\mu_1 - x'\beta) - \Phi\,(-x'\beta) \tag{1333}$$

$$prob\,(y = 2|x) = \Phi\,(\mu_2 - x'\beta) - \Phi\,(\mu_1 - x'\beta) \tag{1334}$$

$$prob\,(y = 3|x) = \Phi\,(\mu_3 - x'\beta) - \Phi\,(\mu_2 - x'\beta) \tag{1335}$$

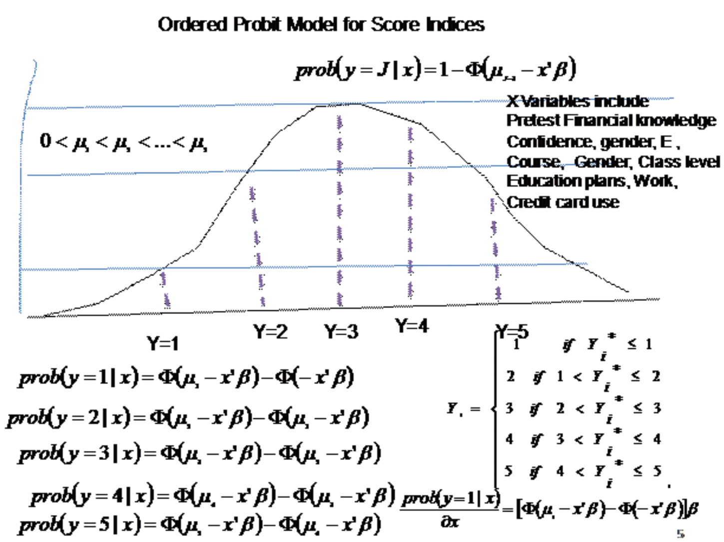
$$prob\,(y = 4|x) = \Phi\,(\mu_4 - x'\beta) - \Phi\,(\mu_3 - x'\beta) \tag{1336}$$

$$prob\,(y = 5|x) = \Phi\,(\mu_5 - x'\beta) - \Phi\,(\mu_4 - x'\beta) \tag{1337}$$

$$0 < \mu_1 < \mu_2 < \ldots < \mu_5 \tag{1338}$$

$$prob\,(y = J|x) = 1 - \Phi\,\left(\mu_{J-1} - x'\beta\right) \tag{1339}$$

Ordered Probit Model



Ordered Probit Model
Marginal effects

$$\frac{prob\,(y = 0|x)}{\partial x} = -\Phi\,(-x'\beta)\,\beta \tag{1340}$$

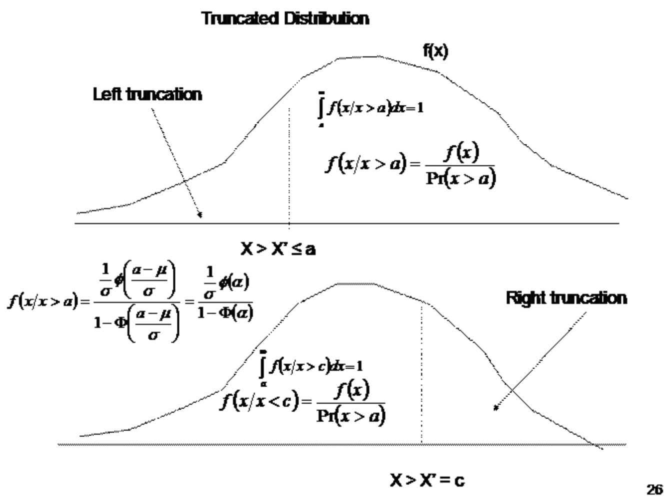
$$\frac{prob\left(y=1|x\right)}{\partial x}=\left[\Phi\left(\mu_1-x'\beta\right)-\Phi\left(-x'\beta\right)\right]\beta \tag{1341}$$

Ordered logit

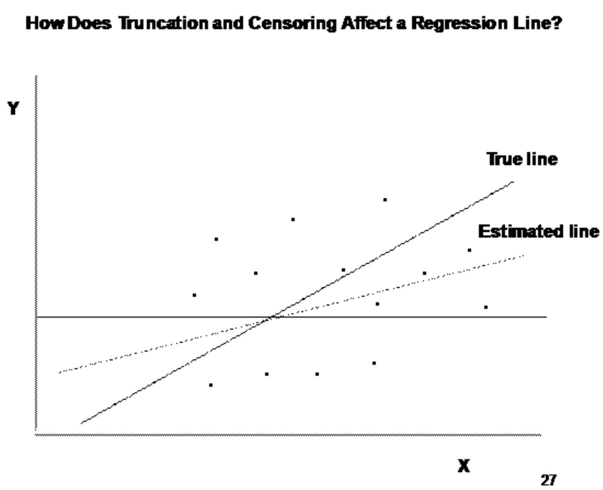
$$Y_{i,t}^{*}=X_{i,t}\beta+\varepsilon_{i,t}. \qquad \varepsilon_{i,t}.|X_{i,t} \sim \ \mathrm{N}\left(0,1\right) \tag{1342}$$

$$Y_{i,t}=j. \quad \text{if} \quad u_{j-1}.<Y_i^{*}\ <\ u_j \qquad \text{j=0,1,...,J} \tag{1343}$$

7.0.20 Truncated Distributions



Censored Data



Mean and Variance For Truncated Distributions
The mean of truncated variable is $E(x/x>a) = \mu + \sigma\lambda(\alpha)$ and
its variance is , $var\,(x/x>a) = \mu + \sigma^2(1-\delta(\alpha))$
where $.0 \leq \delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha] < 1$
Variance is also called a hazard function of this distribution.
$\lambda(\alpha)$ is called inverse Mill's ratio. Here for all $.0 \leq \delta(\alpha) < 1$ for all α.

$$\begin{pmatrix} \lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)} & \text{if x} > \alpha \\ \lambda(\alpha) = \frac{-\phi(\alpha)}{1-\Phi(\alpha)} & \text{if x} < \alpha \end{pmatrix} \tag{1344}$$

Regression with Truncated Distribution
$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ where $\varepsilon_i \sim N(0,\sigma^2)$ $E(Y_i/X_i) = X_i\beta$ $E(Y_i/X_i) \sim N(X_i\beta, \sigma^2)$
if truncated $(Y_i/y_i > \alpha) = X_i\beta + \sigma\lambda(\alpha_i)$

$$\begin{aligned} (Y_i/y_i > \alpha) &= X_i\beta + E(\varepsilon_i/\varepsilon_i > \alpha) > \\ a - X_i\beta &= X_i\beta + \sigma\left(\frac{\frac{\alpha - X_i\beta}{\sigma}}{1-\Phi\left(\frac{\alpha - X_i\beta}{\sigma}\right)}\right) = X_i\beta + \sigma\lambda(\alpha_i) \end{aligned} \tag{1345}$$

This is a non-linear function and the estimates of and by the OLS technique is neither efficient nor consistent. Because the OLS ignores the part;

$$\lambda(\alpha) = \frac{1}{\sigma}\frac{\left(\frac{\alpha - X_i\beta}{\sigma}\right)}{1-\Phi\left(\frac{\alpha - X_i\beta}{\sigma}\right)} = \frac{1}{\sigma}\frac{\frac{1}{\sigma}\phi(\alpha)}{1-\Phi(\alpha)} \tag{1346}$$

$\lambda(\alpha)$ obviously depends on .

Maximum Likelihood Estimator for Truncated Distribution The unknown parameters β and σ can be estimated consistently by
the Maximum likelihood technique as:

$$\begin{aligned} \log L(\alpha,\beta,\sigma) &= -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{N}{2}\log\left(\sigma^2\right) - \\ &\quad \frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{(y_i - \beta X_i)^2}{\sigma^2} - \sum_{i=1}^{N}\left[1-\Phi\left(\frac{(\alpha - \beta X_i)}{\sigma}\right)\right] \end{aligned} \tag{1347}$$

First order conditions for estimation of this ML are:
$\frac{\log L}{\partial\begin{pmatrix}\beta \\ \sigma^2\end{pmatrix}} = \sum_{i=1}^{N}\left[\left(\frac{(y_i-\beta X_i)^2}{\sigma^2} - \frac{\lambda(\alpha_i)}{\sigma}\right)X_i - \frac{1}{2\sigma^2} + \frac{(y_i-\beta X_i)^2}{\sigma^2} - \frac{\lambda(\alpha_i)}{\sigma}\right] = \sum_{i=1}^{N} g_i = 0$
Maximise this likelihood by choosing β and σ iteratively
taking the OLS estimates as the initial starting point.
Censored Regression Model

$$Y_i = \int \begin{array}{l} Y_i^* \;\; if \;\; Y_i^* > 0 \quad \text{if the event occurs} \\ 0 \;\; \text{if } Y_i^* \leq 0 \end{array} \tag{1348}$$

If $Y_i^* \sim N\left(X_i\beta, \sigma^2\right)$ $pr\left(Y_i^* = 0\right) = pr\left(Y_i^* \leq 0\right) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$
Thus the distribution of the transformed variable is

$$g\left(Y_i^*\right) = \left(\begin{array}{ll} N\left(\mu, \sigma^2\right) & \text{if } Y_i^* > 0 \\ 1 - \Phi\left(\frac{\mu}{\sigma}\right) & \text{if } Y_i^* \leq 0 \end{array}\right) \tag{1349}$$

Moments of the censored variable $E\left(Y_i\right) = \Phi\alpha + (1 - \Phi)(\mu + \sigma\lambda)$
and ; $var\left(Y_i\right) = \sigma^2 (1 - \Phi)\left[(1 - \delta) + (\alpha - \lambda)^2\right]\Phi$
$\delta = \lambda^2 - \lambda\alpha$ $\lambda = \frac{\phi}{1-\Phi}$ $E\left(Y_i^* \leq 0\right) = \Phi(\alpha) = \Phi$

7.0.21 Tobit Model

- – It is an extension of the probit model, named after Tobin. We observe variables if the event occurs: ie if some one buys a house. We do not observe explanatory variables for people who have not bought a house. The observed sample is censored, contains observations for only those who buy the house.

$$Y_i = \int \begin{array}{l} Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \quad \text{if the event occurs} \\ 0 = \text{otherwise} \end{array} \tag{1350}$$

  – Y_i is equal to $\beta_1 + \beta_2 X_i + \varepsilon_i$ is the event is observed equal to zero if the event is not observed.

  – It is unscientific to estimate probability only with observed sample without worrying about the remaining observations in the truncated distribution. The Tobit model tries to correct this bias.

  – Inverse Mill's ratio: Example first estimate probability of work then estimate the hourly wage as a function of socio-economic background variables

Tobit Model

$$Y_i = \int \begin{array}{l} Y_i^* \;\; if \;\; Y_i^* > 0 \quad \text{if the event occurs} \\ 0 \;\; \text{if } Y_i^* \leq 0 \end{array} \tag{1351}$$

The unknown parameters and can be estimated consistently by the Maximum likelihood technique as:

$$\log L(\alpha, \beta, \sigma) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) + \log\left(\sigma^2\right) - \frac{1}{2\sigma_1^2}\underbrace{\sum_{i=1}^{N}\left(Y_{1,i} - \beta X_{1,i}\right)^2}_{uncensored} + \underbrace{\sum\left(1 - \Phi_i\right)\left(\frac{\left(X_{2i}\beta - X_{1i}\beta\right)}{\sigma}\right)}_{censored}$$

Uncensored part Censored part
Use the OLS estimates as the starting values.
The two step estimation procedure proceeds as following:
First construct the index variable $I_i = \int \begin{array}{l} Y_i^* \;\; if \;\; Y_i^* > 0 \\ 0 \;\; \text{if } Y_i^* \leq 0 \end{array}$
$pr\left(I_i = 1\right) = 1 - \Phi(\Delta) = p_i$ and $pr\left(I_i = 0\right) = \Phi_i = 1 - p_i$

Then the log likelihood

$$L = \underset{I_i.=1}{\Pi} P_i \underset{I_i.=0}{\overset{\sigma}{\Pi}} (1 - P_i) \tag{1352}$$

Using Probit and MLE find estimates of parameters β and σ and λ $\lambda = \frac{\phi\left(\frac{\beta X_i}{\sigma}\right)}{\Phi\left(\frac{\beta X_i}{\sigma}\right)}$

Secondly apply OLS to $Y_i = \beta_1 + \beta_2 X_i + \sigma\lambda + w_i$

Two Limit Tobit Model $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$

$$Y_i = \begin{pmatrix} Y_i^* \;\; if \;\; L_1 < \; Y_i^* < L_2 \text{ if the event occurs} \\ L_1 \;\; \text{if } Y_i^* < L_1 \\ L_2 \;\; \text{if } Y_i^* \geq L_1 \end{pmatrix} \tag{1353}$$

Formulate a maximum likelihood function

$$L = \Pi\Phi \underset{Y_i^* < L_1}{\left(\frac{L_1 - \; X_i\beta}{\sigma} \right)} \Pi\Phi \underset{Y_i = Y_i^*}{\left(\frac{Y_i - \; X_i\beta}{\sigma} \right)} \Pi\Phi \underset{Y_i^* \geq L_2}{\left(\frac{L_2 - \; X_i\beta}{\sigma} \right)} \tag{1354}$$

For instance think of a minimum, partial or maximum coverage of insurance.
Two Limit Tobit Model
For instance think of a minimum, partial or maximum coverage of insurance.
Expected values

$$E\left(Y_i/X_i, L_1\right) < Y_i^* < L_2 = X_i\beta + \frac{\sigma\left(\Phi_{1,i} - \Phi_{2,i}\right)}{\Phi_{2,i} - \Phi_{1,i}} \tag{1355}$$

$$E\left(Y_i/X_i, L_1\right) = \Phi_{1,i} L_{1,i} + X_i\beta\left(\Phi_{1,i} - \Phi_{2,i}\right) + \sigma\left(\phi_{1,i} - \phi_{2,i}\right) + \left(1 - \Phi_{2,i}\right) L_{2,i} \tag{1356}$$

$$\Phi_{1,i} = \Phi\left(\frac{L_1 - \; X_i\beta}{\sigma}\right); \; \Phi_{2i} = \Phi\left(\frac{L_2 - \; X_i\beta}{\sigma}\right) \tag{1357}$$

$$L = \underset{n_1}{\Pi\Phi_{1,i}} \underset{n_2}{\Pi(\Phi_{2,i} - \Phi_{1,i})} \underset{n_3}{\Pi(1 - \Phi_{2i})} \tag{1358}$$

Estimation of Two Limit Tobit Model

$$L = \underset{n_1}{\Pi\Phi_{1,i}} \underset{n_2}{\Pi(\Phi_{2,i} - \Phi_{1,i})} \underset{n_3}{\Pi(1 - \Phi_{2i})} \tag{1359}$$

$$\log L = N_2 \log\rho + N \log\sigma + \sum_{i=1}^{N} \left(Y_i - X_i\beta\right)^2 - \sum \log\left[\rho\left(1 - \rho\right)\Phi_i\right] \tag{1360}$$

If ρ known estimate β_{ML} and $\sigma._{ML}$ if not β_{ML} and $\sigma._{ML}$ and test for $\rho = 1$

$$E\left(Y_i/X_i;\; I_i = 1\right) = X_i\beta + E\left(\varepsilon_i/I_i = 1\right) = X_i\beta + \sigma\frac{\phi_i}{1 - \Phi_i} \tag{1361}$$

$$E\left(Y_i/X_i;\ I_i=0\right)=X_i\beta+E\left(\varepsilon_i/I_i=0\right)=X_i\beta+\sigma\frac{\phi_i}{1-\Phi_i} \tag{1362}$$

$$E\left(Y_i/X_i\right)=X_i\beta=\frac{\sigma\rho_1\Phi_i-\rho_2\Phi_i}{\rho_1\Phi_i-\rho_2\Phi_i} \tag{1363}$$

7.1 Heckman's Selectivity Bias

Inference population when the sample is non-random and some observations are omitted causes a sample selection bias.

Heckman's procedure is to microeconometrics as is the unit root for time series data.

When the sample selection is not corrected inference drawn from the regression analysis is not efficient or robust.

$Y_{1,i}=X_{1i}\beta+\varepsilon_{1,i}$ if the event occurs if $Y_{1,i}>Y_{2,i}$

$Y_{2,i}=X_{2,i}\beta+\varepsilon_{2,i}$

Both $Y_{1,i}$ and $Y_{2,i}$ are stochastic. For instance, if $Y_{1,i}$ is market wage and $Y_{2,i}$ is the reservation wage; an individual works only when $Y_{1,i}>Y_{2,i}$.

$\begin{pmatrix}\varepsilon_{1,i}\\ \varepsilon_{2,i}\end{pmatrix}\sim IN\begin{pmatrix}0 & \sum\end{pmatrix}\quad \sum=\begin{pmatrix}\sigma_1^2 & \sigma_{1,2}\\ \sigma_{2,1} & \sigma_2^2\end{pmatrix}$

At least one more variable in $X_{2,i}$ than in $X_{1,i}$; if $\sigma_{1,2}=0$ it is a regular Tobit.

But the sample selection problem arise when $\sigma_{1,2}\neq 0$.

$Y_{1,i}>Y_{2,i}$ implies $\varepsilon_{1,i}=Y_{1,i}-X_{1i}\beta$

$\varepsilon_{1,i}<Y_{2,i}-X_{2i}\beta$

Joint t density of $\begin{pmatrix}\varepsilon_{1,i} & \varepsilon_{2,i}\end{pmatrix}\sim IN\begin{pmatrix}0 & \sum\end{pmatrix}=f\begin{pmatrix}\varepsilon_{1,i} & \varepsilon_{2,i}\end{pmatrix}=g\left(\varepsilon_{1,i}\right)h\left(\varepsilon_{2,i}/\varepsilon_{1,i}\right)$

Likelihood Function and Correlated Errors in Heckman's Model

For some individuals you do not observe $Y_{1,i}$ because $Y_{1,i}<Y_{2,i}$.

$Y_{1,i}<Y_{2,i}.{=}\varepsilon_{1,i}-\varepsilon_{2,i}<X_{2i}\beta-X_{1i}\beta$

$\begin{pmatrix}\varepsilon_{1,i} & \varepsilon_{2,i}\end{pmatrix}\sim IN\begin{pmatrix}0 & \sigma^2\end{pmatrix};\ \sigma^2=\sigma_1^2+\sigma_2^2-2\sigma_{1,2}$

$$\begin{aligned}\log L\left(\alpha,\beta,\sigma\right) &= -\frac{N}{2}\log\left(\sigma^2\right)-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N}\underset{obs}{\left(Y_{1,i}-\beta X_{1,i}\right)^2}\\ &\quad +\sum_{Inst}\log\Phi_i\left(w\right)+\sum_{i=1}^{N}\log\left[\underset{unobs}{\left(\frac{\left(X_{2i}\beta-X_{1i}\beta\right)}{\sigma}\right)}\right]\end{aligned} \tag{1364}$$

$$w=\frac{\left(Y_{2,i}-X_{2i}\beta_2\right)}{\sigma_{2,1}}-\frac{\sigma_{1,2}}{\sigma_1^2}\left(\left(Y_{1,i}-X_{1,i}\beta_1\right)\right) \tag{1365}$$

Likelihood Function and Correlated Errors in Heckman's Model

For some individuals you do not observe $Y_{1,i}$ because $Y_{1,i}<Y_{2,i}$.

$$w=\frac{\left(Y_{2,i}-X_{2i}\beta_2\right)}{\sigma_{2,1}}-\frac{\sigma_{1,2}}{\sigma_1^2}\left(\left(Y_{1,i}-X_{1,i}\beta_1\right)\right) \tag{1366}$$

$$W=\beta_0+X_{1,i}\beta_1+\varepsilon_{1,i} \tag{1367}$$

$$W_R = \gamma_0 + \gamma_1 H + \gamma_2 Z + \varepsilon_{2,i} \tag{1368}$$

X and Z are exogenous variables and H and W are endogenous. If $.H = 0 \;\; W_R > W$

$$E(\varepsilon_{2,i}/H > 0) = E(\varepsilon_{2,i}/V > -RD) = E\varepsilon_{2,i}/(\varepsilon_{2,i} - \varepsilon_{1,i}) > \gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1 \tag{1369}$$

$$= \frac{\sigma_2^2 - 2\sigma_{1,2}}{\sigma} \frac{\alpha\phi(\Delta)}{1 - \Phi(\Delta)} = \frac{\alpha\phi(\Delta)}{1 - \Phi(\Delta)} \tag{1370}$$

Heckman's Lamda

$$H = \frac{\beta_0 + X_{1,i}\beta_1 - \gamma_0 - \gamma_2 Z}{\gamma_1} + \frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} \tag{1371}$$

$$W = \begin{pmatrix} \beta_0 + X_{1i}\beta + \varepsilon_{1,i} \\ X\delta + V \end{pmatrix} \text{ if } H > 0 \tag{1372}$$

$$W = H = 0 \quad \text{Otherwise} \tag{1373}$$

$$\Pr(H = 0) = pr(V \leq -RD) = pr\left(\frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} \leq \frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1}{\gamma_1}\right) \tag{1374}$$

$pr(\varepsilon_{2,i} - \varepsilon_{1,i} \leq (\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1)) = \Phi(\Delta)$
$\Delta = \frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1}{\sigma}$; $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}$

$$L = \prod_{H.>0} WH \prod_{H=0}^{\sigma} (\Delta) \tag{1375}$$

$E(\varepsilon_{2,i}/H > 0) = \beta_0 + X_{1,i}\beta_1 + E(\varepsilon_{1,i}/H > 0) = \beta_0 + X_{1,i}\beta_1 + \frac{\alpha\phi(\Delta)}{1-\Phi(\alpha)} = \beta_0 + X_{1,i}\beta_1 + \alpha\lambda$

$$E(\varepsilon_{2,i}/H > 0) = \beta_0 + X_{1,i}\beta_1 + \alpha\lambda = \beta_0 + X_{1,i}\beta_1 + \sigma_{1,2} \frac{\phi\left(\frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1}{\sigma}\right)}{1 - \Phi\left(\frac{\gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1}{\sigma}\right)} \tag{1376}$$

Thus the sample selection bias is due to the Heckman's Lambda term.
Estimation of Heckman's Lamda
To estimate $E(\varepsilon_{2,i}/H > 0) = \beta_0 + X_{1,i}\beta_1 + \alpha\lambda$
Use probit to estimate $\widehat{\lambda} \quad I_i = \begin{pmatrix} 1 & \text{if working} \\ 0 & \text{Otherwise} \end{pmatrix}$
and replace it in $W_i = \beta_0 + X_{1,i}\beta_1 + \alpha\widehat{\lambda}$
$pr(I_i = 1) = pr(H > 0) = 1 - \Phi(\Delta)$
$pr(V > -RD) = pr((\varepsilon_{2,i} - \varepsilon_{1,i}) > \gamma_0 + \gamma_2 Z - \beta_0 - X_{1,i}\beta_1)$; $\quad pr(I_i = 0) = \phi(\Delta)$
Heckman's Tobit then:

$$L = \prod_{H.>0} (1 - \Phi(\Delta)) \prod_{H=0}^{\sigma} \Phi(\Delta) \quad \text{by MLE get and} \tag{1377}$$

by MLE get Δ and $\widehat{\lambda} = \frac{\phi(\Delta)}{1-\Phi(\alpha)}$
$H = \frac{\beta_0 + X_{1,i}\beta_1 - \gamma_0 - \gamma_2 Z}{\gamma_1} + \frac{\varepsilon_{2,i} - \varepsilon_{1,i}}{\gamma_1} = \widehat{\Delta}$
Apply OLS to $W_i = \beta_0 + X_{1,i}\beta_1 + \alpha\widehat{\lambda} + \eta_i$

Estimation of Truncated , Censored and Heckmans' sample selection models Take a cross section dataset as from the BHPS data such as qindresp.sav

There are more than 1400 variables in this data set.
Save data is stata readable format using save as *.dta
Then open the stata. Then set momory 500000 to increase memory.
Then go to statistics/linear models/censored or truncated regression
tobit qprearn qsex qqfachi,11
Go to sample selection
heckman qprearn qsex qqfachi, qage, select(qsex = qprfitb)
Do this practice with smaller dataset. (qfimnl, qfiyr, qfihhmn, qhhsize, qfimnb, qnchild)
LIMDEP is another very useful software for such analysis; see http://people.stern.nyu.edu/wgreene/Hull2013.htm

References

[1] Alvarez J. and M. Arellano (2003) The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators, Econometrica, 71, 4, 1121-1159

[2] Amacher G. S. and D. Hellerstein (1999) The Error Structure of Time Series Cross-Section Hedonic Models with Sporadic Event Timing and Serial Correlation, Journal of Applied Econometrics, 14, 3, 233-252

[3] Andrews DWK (2005) Cross-Section Regression with Common Shocks Econometrica, 73, 5 , 1551-1585

[4] Ashenfelter O C (1986) Handbook of labor economics, North Holland.

[5] Atkinson AB and A. Brandolini (2010) On Analyzing the World Distribution of Income,World Bank Economic Review, 24, 1,1–37

[6] Bai J and S. Ng (2010) Panel Unit Root Tests With Cross-Section Dependence: A Further Investigation Econometric Theory, 26, 4 , 1088-1114

[7] Baltagi B.H. (1984) A Monte Carlo Study for Pooling Time Series of Cross-Section Data in the Simultaneous Equations Model Source: International Economic Review, 25, 3, 603-624

[8] Blundell R (2013) Income Dynamics and the Evolution of Inequality, Royal Economic Society Presidential address, Royal Halloway, University of London, April 4, 2013.

[9] Bhattarai K. (2008) Economic Theory and Models: Derivations, Computations and Applications for Policy Analyses, Serials Publications, New Delhi.

[10] Blundell R W and R J. Smith (1989) Estimation in a class of simultaneous equation limited dependent variable models, Review of Economic Studies, 56:37-38.

[11] Blundell R and I. Walker (1986) A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data, Review of Economic Studies, 53, 4, 539-558

[12] Blundell R W and R J. Smith (1989) Estimation in a class of simultaneous equation limited dependent variable models, Review of Economic Studies, 56:37-38.

[13] Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (2000). The labour market impact of the Working Families Tax Credit', Fiscal Studies, vol. 21(1), pp. 75–103.

[14] Blundell R, L.Pistaferri, and I. Preston (2008) Consumption Inequality and Partial Insurance,American Economic Review, 98:5, 1887–1921

[15] Chesher A (2010) Instrumental Variable Models For Discrete Outcomes, Econometrica, 78, 2, 575–601

[16] Chesher A and A. M. Rosen (2013) What do instrumental variable methods deliver with discrete dependent variables?, American Economic Review: 103, 3 , 557-62

[17] Chesher A, A. M. Rosen and K. Smolinski (2013) An instrumental variable model of multiple discrete choice, Quantitative Economics 4 (2013), 157–196

[18] Chesher A (1984) Improving the efficiency of Probit estimators, Review of Economic Studies, 66:3:523-527.

[19] Coles M. G. and E. Smith (1996) Cross-Section Estimation of the Matching Function: Evidence from England and Wales, Economica, 63, 252, 589-597

[20] Cowing T. G. and A. G. Holtmann (1983) Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data Southern Economic Journal, 49, 3 , 637-653

[21] Chudik, A., Pesaran, M. H. and Tosetti, E. (2011), Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal, 14: C45–C90.

[22] Dobson S and J A Goddard(2011) The Economics of Football (2nd edition), Cambridge University Press.

[23] Fosgerau M, D. McFadden and M.Bierlaire (2013) Choice probability generating functions, Journal of Choice Modelling 8 1–18

[24] Gomes J, L. Kogan, and L. Zhang (2003) Equilibrium Cross Section of Returns, Journal of Political Economy, 111, 4, 693-732

[25] Greene W. (2008) Econometric Analysis, Sixth Edition,Pearson Education, New York.

[26] Greene W.H. (1998) LIMDEP and NLogit Version 11: User Manual, Econometric Software Inc.

[27] Güell M and L. Hu (2006) Estimating the probability of leaving unemployment using uncompleted spells from repeated cross-section data; Journal of Econometrics, 133, 1, 307-341

[28] Hausman J.A., (1978), "Specification Tests in Econometrics", Econometrica, 46, 6, 1251-1271.

[29] Heckman J. J., (1979), "Sample Selection Bias as a Specification Error", Econometrica, 47, 1, pp153-161.

[30] Hey J. D.and C. Orme (1994) Investigating Generalizations of Expected Utility Theory Using Experimental Data, Econometrica, 62, 6, 1291-1326.

[31] Higson C., S. Holly, P. Kattuman and S. Platis (2004) The Business Cycle, Macroeconomic Shocks and the Cross-Section: The Growth of UK Quoted Companies, Economica, 71, 282, 299-318

[32] Imbens G. W. and T Lancaster (1994) Combining Micro and Macro Data in Microeconometric Models, *Review of Economic Studies*, 61:4:655-680.

[33] Judge G.G., W.E. Griffiths, R. C. Hill, H. Lutkepohl and T.C. Lee(1990) Theory and Practice of Econometrics, Wiley.

[34] Kapetanios, G. (2008), A bootstrap procedure for panel data sets with many cross-sectional units. The Econometrics Journal, 11: 377–395.

[35] Keifer N (1988) Economic duration data and hazard functions, *Journal of Economic Literature*, 26:647-679.

[36] Lancaster T and A Chesher (1983) The Estimation of Models of Labour Market Behviour, *Review of Economic Studies*, 50:4:609-624.

[37] Maddala G.S. (1983) Limited-dependent and qualitative variables in econometrics, Cambridge University Press, New York.

[38] McFadden D.L. (1974) Conditional logit analysis of qualitative choice behaviour in P.Zarembka (ed.) Frontiers in Econometrics, New York, Academic Press.

[39] Madsen, E. (2005), Estimating cointegrating relations from a cross section. The Econometrics Journal, 8: 380–405

[40] Moffitt R. (1993) Identification and estimation of dynamic models with a time series of repeated cross-sections, Journal of Econometrics, 59, 1–2, 99-123

[41] Nuamah NNNN(1992) The Error and Covariance Structures of the Mean Approach Model of Pooled Cross-Section and Time Series Data Journal of the Royal Statistical Society,41, 2, 197-207

[42] Nijman T and M. Verbeek (1990) Estimation of time-dependent parameters in linear models using cross-sections, panels, or both Journal of Econometrics, 46, 3, 333-346

[43] Perali F. and J. Chavas (2000) Estimation of Censored Demand Equations from Large Cross-Section Data, American Journal of Agricultural Economics, 82, 4, 1022-1037

[44] Pesaran, . M. . H. , Ullah, A. and Yamagata, T. (2008), A bias-adjusted LM test of error cross-section independence. Econometrics Journal, 11: 105–127

[45] Phillips, P. C. B. and Sul, D. (2003), Dynamic panel estimation and homogeneity testing under cross section dependence, Econometrics Journal, 6: 217–259.

[46] Phillips P.C.B. and D. Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, Journal of Econometrics, 137, 1, 162-188

[47] Theil H. (1969) A multinomial extension of the linear logit model, *International Economic Review*, 10:251-9.

[48] Tobin J (1958) Estimation of relationship for limited dependent variables, *Econometrica*, 26:24-36.

[49] Stoker T. M. (1986) Aggregation, Efficiency, and Cross-Section Regression Econometrica, 54, 1, ,171-188

[50] Staigler D., Stock J. H., (1997), Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 3, 557-586.

[51] White H (1980) Nonlinear Regression on Cross-Section Data, Econometrica, 48, 3 , 721-746

[52] Verbeek M. (2012) A Guide to Modern Econometrics, Wiley, 3rd ed.

[53] Verbeek M and F Vella (2005) Estimating dynamic models from repeated cross-sections, Journal of Econometrics,127, 1, 83-102

7.1.1 Tutorial 4: Probability Models

Q1. Discuss the maximum likelihood functions and Newton-Ralphson or BHHH algorithms for estimation of parameters and the testing procedure for the following cross section models:

a. Logit $\left[P_i = \frac{1}{1+e^{-Z_i}} \text{ with } Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i\right]$.

b. Count data $\left[P\left(Y = y\right) = \frac{e^{-\lambda}\lambda^y}{y!}\right]$.

c. Multinomial Choice model: $\left[\frac{P_{i,2}}{P_{i,1}} = \frac{\frac{\exp\left(X'_{i,2}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}}{\frac{\exp\left(X'_{i,1}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}} = \frac{\exp\left(X'_{i,2}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)}\right]$.

d. Ordered probit model: $\left[prob\left(y = J|x\right) = 1 - \Phi\left(\mu_{J-1} - x'\beta\right)\right]$.

e. Heckman's correction for selectivity bias in which

$Y_{1,i} = X_{1i}\beta + \varepsilon_{1,i}$ and $Y_{2,i} = X_{2,i}\beta + \varepsilon_{2,i}$ and if the event occurs $Y_{1,i} > Y_{2,i}$.

f. Two limit Tobit for a certain regression $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ with

$$Y_i = \left(\begin{array}{l} Y_i^* \ \ if \ \ L_1 < \ Y_i^* < L_2 \text{ if the event occurs} \\ L_1 \ \text{ if } Y_i^* < L_1 \\ L_2 \ \text{ if } Y_i^* \geq L_2 \end{array}\right).$$

8 L7: Panel Data Model

Many economic issues require cause-effect analyses of cross-sections of individuals, households or countries over time. The major issue that economists like to know remains whether coefficients vary across individual observations at a particular time or whether variables have any systematic pattern over time. For instance macroeconomist are interested to know what makes growth rates differ across countries at a particular year and of the same country over time. In other words they want to find out whether there are any country specific and time specific effects on economic growth. Panel studies of growth studies are carried out to know the determinants of growth of an individual country or a group of countries over time. Similarly microeconomic studies aim to investigate whether wages and earnings linked to characteristic of workers and other environmental factors over time or whether profits vary systematically by firms and by production periods.

When confronted with these questions an appropriate econometric method requires using all observations across individuals for each time period under investigation. The paned data regression models in econometrics takes account of all cross section and time series observations. The major emphasis lies on decomposing total variation within a group and between the various groups. Subscript i, t refer to individual and time period respectively.

Literature on the panel data model is developing very fast for instance, Wallace and Hussain (1969), Balestra and Nerlove (1966), Hausman (1978), Chamberlain (1984), Arulampalam and Booth(1998), Blundell and Smith (1989),Chesher (1984), Hansen (1982), Hausman (1978), Heckman (1979), Im, Pesaran and Shin (2003), Imbens and Lancaster (1994), Keifer (1988), Kao (1999), Kwaitkowski, Phillips, Schmidt and Shin (1992), Larsson, Lyhagen and Lothgren (2001) Levin, Lin and Chu (2002), Pedroni (1999), Pesaran and Smith (1995) Phillips (1987), McCoskey and Kao (1999), Johansen Soren (1988), Johansen Soren (1988) Staigler Stock (1997), Lancaster (1979) Lancaster and Chesher (1983) Zellner A. (1985), Weidmeijer (2005). Similarly there are number of excellent texts Baltagi (1995), Davidson R and MacKinnon J. G. (2004) ,Greene W. (2000), Hsiao Cheng (1993), Lancaster (1990), Ruud (2000) Verbeek (2004), Wooldridge (2002). These studies could be grouped as:

- Zellner (1962), Wallace and Hussain (1969), Balestra and Nerlove (1966), Hausman (1978), Chamberlain (1984), Arulampalam and Booth(1998), Blundell and Smith (1989),Chesher (1984), Hansen (1982), Hausman (1978), Heckman (1979), Im, Pesaran and Shin (2003), Imbens and Lancaster (1994), Keifer (1988), Kao (1999), Kwaitkowski, Phillips, Schmidt and Shin (1992), Larsson, Lyhagen and Lothgren (2001) Levin, Lin and Chu (2002), Pedroni (1999), Pesaran and Smith (1995), Phillips (1987), McCoskey and Kao (1999), Staigler and Stock (1997), Lancaster (1979), Lancaster and Chesher (1983), Zellner (1985), Weidmeijer (2005).

- Nickell (1981), Arellano and Bond (1991), Arellano and Bover (1995), Kiviet (1995), Islam (1995), Mankiew, Romer and Weil (1992), Caselli, Esquivel, Lefort (1996), Blundell and Bond (1998), Judson and Owen (1999), Hahn and Kuersteiner (2002), Ho (2006), Windmeijer (2005), Roodman (2009), Wooldridge (2010).

- Hansen(1999), Kleibergen and Paap (2006), Hayakawa(2009), Baltagi and Feng, Kao (2012), Su and Lu (2013), Kapetanios, Mitchell and Shin (2014), Lee (2014)

- Cornwell and Rupert (1988), Robertson and Symons (1992), Hansen (1999), Bai and Ng (2005). Wooldridge (2005), Kleibergen and Paap (2006), Sentana(2009), Carriero, Kapetanios and Marcellino (2009), Semykina and Wooldridge (2013), Koop (2013)

- Similarly there are number of excellent texts Johnston (1960), Baltagi (1995), Davidson and MacKinnon (2004), Greene (2000), Hsiao (1993), Lancaster (1990), Ruud (2000), Verbeek (2004), Wooldridge (2002).

8.0.2 Structure of Panel Data

for $i = 1, \ldots . N$ countries and $t = 1, \ldots ., T$ years

Table 20: Structure of Panel Data

Dependent Variable	Explanatory Variable	Random Error
$y_{1,1}$	$x_{1,1}$	$e_{1,1}$
.	.	.
$y_{1,T}$	$x_{1,T}$	$e_{1,T}$
$y_{2,1}$	$x_{2,1}$	$e_{2,1}$
.	.	.
$y_{2,T}$	$x_{2,T}$	$e_{2,T}$
.	.	.
$y_{N,1}$	$x_{N,1}$	$e_{,1}$
.	.	.
$y_{2,T}$	$x_{2,T}$	$e_{2,T}$

8.1 Advantages of Panel Data Model

- Large number of observations over individuals and time make estimates more efficient and asymptotically consistent
- Possible to check individual and time effects in a regression
- Very inclusive and comprehensive, state and space dimensions
- Can use vast amount of information from census, household surveys, firm or country wise statistics
- Background for testing economic theories at micro as well as macro level

Recent literature on Panel Data Model

- Theory of Panel Data Estimation
- Pooling time sereis and cross section: SUR
- Between and Within Effects
- Fixed and Random Effect Models
- Dynamic Panel

- Panel Unit Root and Panel Cointegration
- Panel Data Model for Limited Dependent Variables

Lessons from Static and Dynamic Panel data Models
Economic Growth of Countries around the World:
Unemployment-inflation Trade-offs in OECD Countries

8.2 Pooling Cross Section and Time Series: Seemming Unrelated Regression (SUR) MODEL

A SURE model includes m endogenous variables and a system of m equations. It may have other exogenous variables. For example, in spirit of the Phillips curve analysis one may think that growth rate of output in period t may be influenced positively by the inflation in the last period, t-1. The reason is obvious. At the given capacity to produce, higher prices may be indicative of the higher level of final demand. Higher demand causes a rise in prices of goods in response to higher factor prices necessary to induce overtime work by workers or extra shifts of machine hours in the production process. One may argue therefore that producer supply more today only if prices of goods were higher in the last period. Alternative hypothesis, might be that growth rate has negative relation with inflation, as inflation creates uncertainty and harms investment. Rational expectation school rules out any systematic impact of prices on output. Which one of these claims are true? This is a question of empirical nature.

- SUR if formed by stacking models

$$Y_1 = X_1\beta + e_1 \tag{1378}$$

$$Y_2 = X_2\beta + e_2 \tag{1379}$$

$$... \tag{1380}$$

$$Y_m = X_m\beta + e_m \tag{1381}$$

There are m equations and T observations in the SURE system (in growth rate example we have 151 countries and 31 observations). They can be stacked into one large equation system as following.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & . & . & 0 \\ 0 & X_2 & . & . & 0 \\ . & . & X_3 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & X_m \end{bmatrix} \begin{matrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_m \end{matrix} + \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_m \end{bmatrix} \tag{1382}$$

- Each Y_m and $e_m$has a dimension of T by 1 and X_m has T by K dimension and each β_m has K by 1 dimension. The covariance matrix of errors has TM by TM dimension.

Seemming Unrelated Regression (SUR) MODEL: Assumptions

- Mean of $e_{i,t}$ is zero for every value of , $E(e_{i,t}) = 0$
- variance of $e_{i,t}$ is constant for every ith observation, $var(e_{1t}) = \sigma_i^2$
- $cov(e_{i,t}, e_{i,s}) = 0$ for al t=s; this also means there is no autocorrelation
- All of the above assumptions are standard to the OLS assumptions.
- The major difference lies on assumption of contemporaneous correlation across the disturbance terms in above two models.
- $cov(e_{i,t}, e_{j,s}) = \sigma_{i,j}^2$ The systems are related due to errors.

Variance Covariance Structure in SUR MODEL

- Dimension of each of the $\sigma_{i,j}$, like that of the identity matrix I, is T by T, and reflects the variance covariance matrix of the stacked regression.
- The Kronnecker product $\Sigma \otimes I$is a short way of writing this covariance matrix.
- Σ is the variance covariance matrix
- $\otimes$ is the symbol for the Kronnecker product
- I is Identity Matrix with T×M by T×M dimension.

Pooling Cross Section and Time Series: Seemming Unrelated Regression (SUR) MODEL

- $$ee\prime = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_m \end{bmatrix} \begin{bmatrix} e_1 & e_2 & . & . & e_m \end{bmatrix} = \begin{bmatrix} e_1^2 & e_1e_2 & e_1e_3 & e_1e_4 & e_1e_5 \\ e_1e_2 & e_2^2 & e_2e_3 & e_2e_4 & e_2e_5 \\ e_1e_3 & e_2e_3 & e_3^2 & e_3e_4 & e_4e_5 \\ e_1e_4 & e_4e_2 & e_4e_3 & e_4^2 & e_4e_5 \\ e_1e_5 & e_5e_2 & e_5e_3 & e_5e_4 & e_5^2 \end{bmatrix} \tag{1383}$$

$$E(ee\prime) = \begin{bmatrix} var(e_1) & cov(e_1e_2) & cov(e_1e_3) & cov(e_1e_4) & cov(e_1e_5) \\ cov(e_1e_2) & var(e_2) & cov(e_2e_3) & cov(e_2e_4) & cov(e_2e_5) \\ cov(e_1e_3) & cov(e_2e_3) & var(e_3) & cov(e_3e_4) & cov(e_4e_5) \\ cov(e_1e_4) & cov(e_4e_2) & cov(e_4e_3) & var(e4) & cov(e_4e_5) \\ cov(e_1e_5) & cov(e_5e_2) & cov(e_5e_3) & cov(e_5e_4) & var(e_5) \end{bmatrix} \tag{1384}$$

Pooling Cross Section and Time Series: Seemming Unrelated Regression (SUR) MODEL

- $$E(ee\prime) = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \sigma_{1,4} & \sigma_{1,5} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & \sigma_{2,4} & \sigma_{2,5} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} & \sigma_{3,5} \\ \sigma_{4,1} & \sigma_{4,2} & \sigma_{4,3} & \sigma_4^2 & \sigma_{4,5} \\ \sigma_{5,1} & \sigma_{5,2} & \sigma_{5,3} & \sigma_{5,4} & \sigma_5^2 \end{bmatrix} = V = \Sigma \otimes I \tag{1385}$$

Application of the OLS technique individual equations generates inconsistent results. Sure method aims to correct this problem by estimating all equations simultaneously.

The SURE method is essentially a generalised least square estimator. Note

$$V^{-1} = \Sigma^{-1} \otimes I \tag{1386}$$

Pooling Cross Section and Time Series: Seemming Unrelated Regression (SUR) MODEL
Aitken generalised least square

$$\widehat{\beta} = \left[X'V^{-1}X\right]^{-1} X'V^{-1}Y = \left[X'\left(\Sigma^{-1} \otimes I\right)X\right]^{-1} X'\left(\Sigma^{-1} \otimes I\right)Y \tag{1387}$$

$$\widehat{\beta} = \begin{bmatrix} \sigma_{1,1}X_1'X_1 & \sigma_{1,1}X_1'X_2 & \sigma_{1,1}X_1'X_3 & \sigma_{1,m}X_1'X_m \\ \sigma_{2,1}X_2'X_1 & \sigma_{2,2}X_2'X_2 & \sigma_{2,3}X_2'X_3 & \sigma_{2,m}X_2'X_m \\ & & & \\ & & & \\ \sigma_{m,1}X_m'X_1 & \sigma_{m,2}X_m'X_2 & \sigma_{m,3}X_m'X_3 & \sigma_{m,4}X_m'X_m \end{bmatrix} \begin{bmatrix} \sum \sigma_{1,j}X_1'Y_j \\ \\ \\ \\ \sum \sigma_{m,j}X_m'Y_j \end{bmatrix} \tag{1388}$$

Steps for SUR Estimation

- Estimate each equation separately using the least square technique.
- Use the least square residuals from step 1 to estimate the error term.
- Use the estimates from the second step to estimate two equations jointly within a generalised least square framework. If m=2 the variance covariance matrix will be as given below.

Estimation of Seemming Unrelated Regression (SUR) by GLS

- $$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{pmatrix} \tag{1389}$$

  Using a theorem in matrix algebra W can be decomposed into two parts as

$$P'P = \Omega^{-1} \tag{1390}$$

Use this partition of Ω to transform the original model as

$$Y = X\beta + \varepsilon \tag{1391}$$

$$\beta_{OLS} = \left(X'X\right)^{-1}\left(X'Y\right) \tag{1392}$$

Estimation of Seemming Unrelated Regression (SUR) by GLS
Transform it to

$$P'Y = P'X\beta + P'\varepsilon \tag{1393}$$

$$Y^* = X^*\beta + \varepsilon^* \tag{1394}$$

$$\beta_{GLS} = (X'P'PX)^{-1}(X'P'PY) \tag{1395}$$

In matrix notation

$$\beta_{GLS} = \left(X^{*\prime}\Omega^{-1}X^*\right)^{-1}\left(X^{*\prime}\Omega^{-1}Y^*\right) \tag{1396}$$

Ω^{-1} is inverse of variance covariance matrix.

The GLS estimates are best, linear and unbiased estimators of the coefficients in the SURE system.

8.3 Total, within and between group estimates

Total Effect (pooled model)

$$\beta_{OLS} = \frac{\sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}\right)\left(Y_{i,t}-\overline{Y}\right)}{\sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}\right)\left(X_{i,t}-\overline{X}\right)} = \frac{t_{x,y}}{t_{x,x}} \tag{1397}$$

Here $t_{x,x} = \sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}\right)\left(X_{i,t}-\overline{X}\right)$ and $t_{x,y} = \sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}\right)\left(Y_{i,t}-\overline{Y}\right) = W_{x,y} + b_{x,y}$.

$$t_{x,y} = \sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}_i+\overline{X}_i-\overline{X}\right)\left(Y_{i,t}-\overline{Y}_i+\overline{Y}_i-\overline{Y}\right) \tag{1398}$$

$$t_{x,y} = \sum_t^T\sum_i^N\left(X_{i,t}-\overline{X}_i\right)\left(Y_{i,t}-\overline{Y}_i\right) + T\sum_i^N\left(\overline{X}_i-\overline{X}\right)\left(\overline{Y}_i-\overline{Y}\right) = W_{x,y} + b_{x,y} \tag{1399}$$

Within estimator or fixed effects estimator (variance around the group means)

$$\beta_w = \frac{W_{x,y}}{W_{x,x}} = \frac{\sum_i^N\sum_t^T\left(X_{i,t}-\overline{X}_i\right)\left(Y_{i,t}-\overline{Y}_i\right)}{\sum_i^N\sum_t^T\left(X_{i,t}-\overline{X}_i\right)^2} \tag{1400}$$

Between group effect (variation of group means around the overall means)

$$\beta_b = \frac{b_{x,y}}{b_{x,x}} = \frac{\sum_i^N T\left(\overline{X}_i-\overline{\overline{X}}\right)\left(\overline{Y}_i-\overline{\overline{Y}}\right)}{\sum_i^N T\left(\overline{X}_i-\overline{\overline{X}}\right)^2} \tag{1401}$$

$$\beta_{OLS}t_{x,x} = t_{x,y} = W_{x,y} + b_{x,y} = \beta_W \frac{W_{x,x}}{W_{x,y} + b_{x,y}} + \beta_b \frac{b_{x,x}}{W_{x,y} + b_{x,y}} \tag{1402}$$

Example

Consider the cross-regional variation of expenditure on food in the UK. For simplicity, it is assumed that food expenditure depends only on wage and salary income in each region.

1. Formulate a model relating expenditure on food (F) and income (Y) that takes account of region specific effects. Note that the equations for each region are independent but that there is contemporaneous correlation among the error terms across the regions. State the major assumptions of the model.

2. Represent the model in terms of a system of stacked regressions that takes account of both individual and system specific effects. What is the structure of the covariance matrix of the error terms in this system?

3. Show how the SURE or GLS estimator system can be applied to estimate the structural parameters of this model. Write out their covariance structure in the matrix form.

4. This model has been estimated using a pooled time series and cross section data set (with the sample size of T=14 and N=13) available from the web site of the Office of the National Statistics (hhttp://www.statistics.gov.uk). The estimated coefficients, by region, are given in the following table. Analyse and interpret these results.

8.4 Panel Data: Fixed Effects

$$y_{i,t} = \alpha_i + x_{i,t}\beta + e_{i,t} \qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right) \tag{1403}$$

where parameter α_i picks up the fixed effects that differ among individuals but constant over time,β is the vector of coefficients on explanatory variables. These parameters can be estimated by OLS when N is small but not when that is large.

The model need to be transformed to the least square dummy variable method when N is too large. For this take time average

$$\overline{y}_i = \alpha_i + \overline{x}_i\beta + e_i \qquad \overline{y}_i = T^{-1}\sum_i y_{i,t} \tag{1404}$$

Take the mean difference

$$y_{i,t} - \overline{y}_i = \left(x_{i,t} - \overline{x}_i\right)\beta + \left(e_{i,t} - e_i\ \right) \tag{1405}$$

fixed effect least square dummy variable estimator of β is

$$\beta_{FE} = \left(\sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(x_{i,t} - \overline{x}_i\right)'\right)^{-1} \sum_t^T\sum_i^N \left(x_{i,t} - \overline{x}_i\right)\left(y_{i,t} - \overline{y}_i\right)' \tag{1406}$$

$$\alpha_i = \overline{y}_i - \overline{x}_i\beta_{FE} \tag{1407}$$

Panel Data Model: Fixed Effect

fixed effect least square dummy variable estimator of β is

$$\beta_{FE} = \left(\sum_t^T \sum_i^N (x_{i,t} - \overline{x}_i)(x_{i,t} - \overline{x}_i)' \right)^{-1} \sum_t^T \sum_i^N (x_{i,t} - \overline{x}_i)(y_{i,t} - \overline{y}_i)' \tag{1408}$$

$$\alpha_i = \overline{y}_i - \overline{x}_i \beta_{FE} \tag{1409}$$

These estimators are unbiased, consistent and efficient with corresponding covariance matrix given by:

$$cov(\beta_{FE}) = \sigma_e^2 \left(\sum_t^T \sum_i^N (x_{i,t} - \overline{x}_i)(x_{i,t} - \overline{x}_i)' \right)^{-1} \tag{1410}$$

where

$$\sigma_e^2 = \frac{1}{N(T-1)} \sum_t^T \sum_i^N (y_{i,t} - \alpha_i - x_{i,t}\beta_{FE}) \tag{1411}$$

Panel Data Model: Fixed Effect in Matrix Notation

$$y_i = i\alpha_i + \overline{X}_i \beta + e_i \tag{1412}$$

$$\begin{bmatrix} Y_1 \\ Y_1 \\ . \\ . \\ Y_N \end{bmatrix} = \begin{bmatrix} I & 0 & . & . & 0 \\ 0 & I & . & . & 0 \\ . & . & I & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & I \end{bmatrix} \begin{matrix} Y_1 \\ Y_1 \\ . \\ . \\ Y_N \end{matrix} + \begin{bmatrix} X_1 \\ X_1 \\ . \\ . \\ X_N \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ e_1 \\ . \\ . \\ e_N \end{bmatrix} \tag{1413}$$

$$Y = \begin{bmatrix} d_1 & d_2 & . & . & d_N & X \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{1414}$$

$$Y = D\alpha + X\beta + e \tag{1415}$$

Panel Data Model: Fixed Effect in Matrix Notation

This can be easily estimated by the OLS when the number of cross section units are small. Many panel data studies have much larger observations. It results in over parameterisation and loss of degree of freedom. For this the model is transformed by a projection matrix

$$M_d = I - D(D'D)D' \tag{1416}$$

$$M_d Y = ID\alpha - M_d X\beta + M_d \cdot e \tag{1417}$$

$$M_d = \begin{bmatrix} M^0 & 0 & . & 0 & . & 0 \\ . & M^0 & . & 0 & . & 0 \\ . & 0 & M^0 & 0 & . & 0 \\ . & . & . & M^0 & . & . \\ . & 0 & . & 0 & . & 0 \\ . & . & . & . & . & M^0 \end{bmatrix} \; wehre \; M^0 = I_T - \frac{1}{T} ii' \tag{1418}$$

$$M^0 = I_T - \frac{1}{T} ii' \tag{1419}$$

Panel Data Model: Fixed Effect in Matrix Notation
Multiplying any variable by M^0 is equivalent taking deviation from the mean ie

$$M^0 X_i = X_i - \overline{X}_i \tag{1420}$$

$$Y = D\alpha + X\beta + e \tag{1421}$$

$$\alpha = (D'D)^{-1} D (Y - Xb) \tag{1422}$$

$$var(b) = s^2 [X' M_d X]^{-1} \tag{1423}$$

and

$$s^2 = \frac{\sum_t^T \sum_i^N (y_{i,t} - \alpha_i - x_{i,t} b)}{NT - N - K} \tag{1424}$$

8.4.1 Panel Data Model: Random Effect

Random effect models are more appropriate for analysing determinants of growth as

$$y_{i,t} = \mu + x_{i,t}\beta + \alpha_i + e_{i,t} \tag{1425}$$

where $\alpha_i \sim IID\left(0, \sigma_\alpha^2\right)$ are individual specific random errors and $e_{i,t} \sim IID\left(0, \sigma_e^2\right)$ are remaining random errors.

$$\alpha_i \iota_T + e_i \qquad where\ \iota_T = (1, 1, ...1) \tag{1426}$$

$$var(\alpha_i \iota_T + e_i) = \Omega = \sigma_\alpha^2 \iota_T \iota_T' + \sigma_e^2 I_T \tag{1427}$$

Errors are correlated therefore this requires estimation by the Generalised Least Square estimator. Transform the model by pre-multiplying by Ω^{-1} where

$$\Omega^{-1} = \sigma_e^2 \left[I_T - \frac{\sigma_\alpha^2}{\sigma_e^2 + T\sigma_\alpha^2} \iota_T \iota_T' \right] \tag{1428}$$

Panel Data Model: Random Effect

$$\begin{aligned} \beta_{GLS} \quad = \quad & \left(\sum_t^T \sum_i^N (x_{i,t} - \overline{x}_i)(x_{i,t} - \overline{x}_i)' + \psi T \sum_i^N (x_{i,t} - \overline{x}_i)(x_{i,t} - \overline{x}_i)' \right)^{-1} \\ & \left(\sum_t^T \sum_i^N (x_{i,t} - \overline{x}_i)(y_{i,t} - \overline{y}_i)' + \psi T \sum_i^N (x_{i,t} - \overline{x}_i)(y_{i,t} - \overline{y}_i)' \right) \end{aligned} \tag{1429}$$

$$\Omega = \begin{bmatrix} \sigma_\alpha^2 + \sigma_e^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & . & . & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_e^2 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & . & . & \sigma_\alpha^2 + \sigma_e^2 \end{bmatrix} \tag{1430}$$

$$\Omega^{-\frac{1}{2}} = \frac{1}{\sigma_e}\left[I_T - 1 - \frac{\sigma_e}{\sqrt{\sigma_e^2 + T\sigma_\alpha^2}}\right] \tag{1431}$$

$$\beta_{GLS} = \sum_i \left(X'\Omega^{-1}X\right)^{-1} \sum_i \left(X'\Omega^{-1}Y\right) \tag{1432}$$

8.4.2 Dynamic Panel Data Model: GMM Estimator

generalised method of moments (GMM) as proposed by Hansen (1982).

$$y_{i,t} = \gamma y_{i,t-1} + x_{i,t}\beta + \alpha_i + e_{i,t} \qquad \gamma < 1 \tag{1433}$$

which generates the following estimator

$$\gamma_{_{FE}} = \frac{\sum_t^T \sum_i^N \left(y_{i,t} - \overline{y}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right)}{\sum_t^T \sum_i^N \left(y_{i,t} - \overline{y}_{i,t-1}\right)^2}; \qquad \overline{y}_i = T^{-1}\sum_i y_{i,t}; \text{and } \overline{y}_{i,-1} = T^{-1}\sum_i y_{i,t-1} \tag{1434}$$

This is not asymptotically unbiased estimator:

$$\gamma_{_{FE}} = \gamma + \frac{\left(\frac{1}{NT}\right)\sum_t^T \sum_i^N \left(e_{i,t} - \overline{e}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right)}{\left(\frac{1}{NT}\right)\sum_t^T \sum_i^N \left(y_{i,t} - \overline{y}_{i,-1}\right)^2} \tag{1435}$$

$$\underset{N\to\infty}{p\lim} \left(\frac{1}{NT}\right)\sum_t^T \sum_i^N \left(e_{i,t} - \overline{e}_i\right)\left(y_{i,t} - \overline{y}_{i,t-1}\right) = -\frac{\sigma_e^2}{T^2}\frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2} \neq 0 \tag{1436}$$

Panel Data Model: Instrumental Variables for GMM
Instrumental variable methods have been suggested to solve this inconsistency

$$\widehat{\gamma}_{_{IV}} = \frac{\sum_t^T \sum_i^N y_{i,t-2}\left(y_{i,t-1} - \overline{y}_{i,t-2}\right)}{\sum_t^T \sum_i^N y_{i,t-2}\left(y_{i,t-1} - y_{i,t-2}\right)^2} \tag{1437}$$

where $y_{i,t-2}$ is used as instrument of $(y_{i,t-1} - y_{i,t-2})$
It is asymptotically

$$\underset{N\to\infty}{p\lim} \left(\frac{1}{NT}\right) \sum_{t}^{T}\sum_{i}^{N} (e_{i,t} - \overline{e}_i)\, y_{i,t-2} = 0 \tag{1438}$$

Moment conditions with vector of transformed error terms

$$\Delta e_i = \begin{pmatrix} e_{i,2} - e_{i,1} \\ e_{i,3} - e_{i,2} \\ . \\ . \\ e_{i,T} - e_{i,T-1} \end{pmatrix} \tag{1439}$$

Panel Data Model: Instrumental Variables for GMM 2

$$Z_i = \begin{bmatrix} [y_{i,0}] & 0 & . & . & 0 \\ 0 & [y_{i,0}, y_{i,1,}] & 0 & . & . \\ 0 & 0 & . & . & 0 \\ . & . & . & . & 0 \\ 0 & 0 & . & . & [y_{i,0}, y_{i,T-2}] \end{bmatrix} \tag{1440}$$

$$E\left\{Z_i^{'} \Delta e_i\right\} = 0 \tag{1441}$$

Or for moment estimator write the transformed errors as

$$E\left\{Z_i^{'} \left(\Delta y_{i,t} - \gamma \Delta y_{i,t}\right)\right\} = 0 \tag{1442}$$

$$\min_{\gamma} \left(\left(\frac{1}{N}\right) \sum_{i=1}^{N} Z_i^{'} \left(\Delta y_{i,t} - \gamma \Delta y_{i,t}\right)\right)^{'} W_N \sum_{i=1}^{N} Z_i^{'} \left(\Delta y_{i,t} - \gamma \Delta y_{i,t}\right)^{\prime} \tag{1443}$$

Panel Data Model: Instrumental Variables for GMM 2
GMM method includes the most efficient instrument

$$\gamma_{_{GMM}} = \left(\left(\sum_{i=1}^{N} \Delta y_{i,t} Z_i\right) W_N \left(\sum_{i=1}^{N} Z_i^{'} \Delta y_{i,t}\right)\right)^{-1} \times \left(\left(\sum_{i=1}^{N} \Delta y_{i,t} Z_i\right) W_N \left(\sum_{i=1}^{N} Z_i^{'} \Delta y_{i,t}\right)\right) \tag{1444}$$

Blundell and Smith (1989) and Verbeek (2004), Wooldridge (2002) among others have more extensive exposure in GMM estimation. The essence of the GMM estimation remains in finding a weighting matrix that can guarantee the most efficient estimator. This should be inversely proportional to transformed covariance matrix.

$$W_N^{opt} = \left(\left(\frac{1}{N}\right) \sum_{i=1}^{N} Z_i^{'} \Delta e_{i,t} \Delta e_{i,t}^{'} Z_i\right)^{-1} \tag{1445}$$

Panel Data Model: Instrumental Variables for GMM 2

Doornik and Hendry (2001, chap. 7-10) provide a procedure on how to estimate coefficients using fixed effect, random effect and the GMM methods including a lagged terms of dependent variable among explanatory variables for a dynamic panel data model:

$$y_{i,t} = \sum_{i=1}^{p} a_k y_{i,t-s} + \beta^t (L) x_{i,t} + \lambda_t + \alpha_i + e_{i,t} \quad \textit{or in short} \;\; y_{i,t} = W_i \delta + \iota_i a_i + e_i \tag{1446}$$

The GMM estimator with instrument (levels, first differences, orthogonal deviations, deviations from individual means, combination of first differences and levels) used in PcGive is :

$$\widehat{\delta} = \left(\left(\sum_{i=1}^{N} W_i^* Z_i \right) A_N \left(\sum_{i=1}^{N} Z_i' W_i \right) \right)^{-1} \left(\left(\sum_{i=1}^{N} W_i^* Z_i \right) A_N \left(\sum_{i=1}^{N} Z_i' y_i^* \right) \right) \tag{1447}$$

where $A_N = \left(\sum_{i=1}^{N} Z_i' H_i Z_i \right)^{-1}$ is the individual specific weighting matrix.

Hausman Test for Fixed over Random Effect Models

$$\begin{aligned}
var\left(b - \widehat{\beta}\right) &= var\,(b) + var\left(\widehat{\beta}\right) - cov\left(b, \widehat{\beta}\right) - cov\left(\widehat{\beta}, b\right) \\
cov\left[\left(b, \widehat{\beta}\right), \widehat{\beta}\right] &= cov\left[\left(b, \widehat{\beta}\right)\right] - var\left(\widehat{\beta}\right) = 0 \\
cov\left[\left(b, \widehat{\beta}\right)\right] &= var\left(\widehat{\beta}\right) \\
var\left(b, \widehat{\beta}\right) &= var\,(b) - var\left(\widehat{\beta}\right) = \Psi \qquad (1448) \\
W &= \chi^2\,[K-1] = \left(b - \widehat{\beta}\right)' \Psi^{-1} \left(b - \widehat{\beta}\right) \qquad (1449)
\end{aligned}$$

8.4.3 Panel Estimation

Panel Cointegration

8.4.4 Estimation of Panel Data Model with BHPS in SATA

Take a cross section dataset as from the BHPS data such as qindresp.sav

determine panel id: See log file; panel.smcl

Command for

- random effect: xtreg qprearn qsex qqfachi age_years, re
- Fixed effect: xtreg qprearn qsex qqfachi age_years, re
- Between effect: xtreg qprearn qsex qqfachi age_years, be
- MLE:xtreg qprearn qsex qqfachi age_years, mle

These calculations can be done with do command in STATA. For this import analysis.csv datafile and run "do" file file analysis.do.

```
. xtreg qprearn qsex qqfachi age_years, re

Random-effects GLS regression                   Number of obs      =     14910
Group variable: qdoiy4                          Number of groups   =         3

R-sq:  within  = 0.1641                         Obs per group: min =        31
       between = 0.9944                                        avg =    4970.0
       overall = 0.1636                                        max =     14077

Random effects u_i ~ Gaussian                   Wald chi2(3)       =   2915.32
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     qprearn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        qsex |   .3702326   .0105693    35.03   0.000     .3495171    .3909481
     qqfachi |  -.2928379   .0054389   -53.84   0.000    -.3034979   -.2821778
   age_years |   .0731788   .0149741     4.89   0.000     .0438302    .1025274
       _cons |  -7.350651   .0552183  -133.12   0.000    -7.458877   -7.242425
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  1.8103504
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. xtreg qprearn qsex qqfachi age_years, fe

Fixed-effects (within) regression               Number of obs      =     14910
Group variable: qdoiy4                          Number of groups   =         3

R-sq:  within  = 0.1642                         Obs per group: min =        31
       between = 0.9782                                        avg =    4970.0
       overall = 0.1635                                        max =     14077

                                                F(3,14904)         =    975.78
corr(u_i, Xb)  = -0.0392                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     qprearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        qsex |   .3642664   .0107377    33.92   0.000     .3432192    .3853136
     qqfachi |  -.2933891   .0054393   -53.94   0.000    -.3040509   -.2827274
   age_years |   .0692979   .0150113     4.62   0.000     .0398739    .0987219
       _cons |  -7.327959   .0556048  -131.79   0.000    -7.436951   -7.218966
-------------+----------------------------------------------------------------
     sigma_u |  .22423122
     sigma_e |  1.8103504
         rho |  .01510966   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(2, 14904) =     6.12            Prob > F = 0.0022
```

```
. xtreg qprearn qsex qqfachi age_years, mle

Fitting constant-only model:
Iteration 0:   log likelihood = -31468.209
Iteration 1:   log likelihood = -31345.516
Iteration 2:   log likelihood = -31341.428
Iteration 3:   log likelihood = -31341.422

Fitting full model:
Iteration 0:   log likelihood = -30008.463
Iteration 1:   log likelihood = -30007.658
Iteration 2:   log likelihood =   -30006.9
Iteration 3:   log likelihood =  -30006.83
Iteration 4:   log likelihood = -30006.825
Iteration 5:   log likelihood = -30006.825

Random-effects ML regression                    Number of obs      =     14910
Group variable: qdoiy4                          Number of groups   =         3

Random effects u_i ~ Gaussian                   Obs per group: min =        31
                                                               avg =    4970.0
                                                               max =     14077

                                                LR chi2(3)         =   2669.19
Log likelihood  = -30006.825                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     qprearn |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        qsex |   .3651194   .0107613    33.93   0.000     .3440277    .3862111
     qqfachi |  -.2932528   .0054391   -53.92   0.000    -.3039133   -.2825924
   age_years |   .0701004   .0150188     4.67   0.000     .0406641    .0995367
       _cons |  -7.429477   .0960936   -77.31   0.000    -7.617817   -7.241137
-------------+----------------------------------------------------------------
    /sigma_u |   .1016117    .063529                      .0298374    .3460402
    /sigma_e |   1.810166   .0104834                      1.789735     1.83083
         rho |   .0031411   .0039157                       .000202    .0269364
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=    4.11 Prob>=chibar2 = 0.021
```

```
. xtreg qprearn qsex qqfachi age_years, be
note: age_years omitted because of collinearity

Between regression (regression on group means)  Number of obs      =     14910
Group variable: qdoiy4                          Number of groups   =         3

R-sq:  within  = 0.1627                         Obs per group: min =        31
       between = 1.0000                                        avg =    4970.0
       overall = 0.1621                                        max =     14077

                                                F(2,0)             =         .
sd(u_i + avg(e_i.))=         0                  Prob > F           =         .

------------------------------------------------------------------------------
     qprearn |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        qsex |  -.2944859          .        .       .            .           .
     qqfachi |   .2253665          .        .       .            .           .
   age_years |  (omitted)
       _cons |  -8.326359          .        .       .            .           .
------------------------------------------------------------------------------
```

Table 21: Determinants of growth rate of per capita income and Exchange Rate

	Growth Model		Exchange Rate Model	
Determinants	Coefficient	t-prob	Coefficient	t-value
Investment ratio	0.1820	.00060	-	-
Export Ratio	0.0257	.3830	-	-
Exchange rate -1	-	-	0.9710	0.00
Real Interest rate	-	-	-0.0290	0.00
Population growth rate	-0.8849	0.1540	0.7917	0.00
Constant	3.0116	0.1780	0.3400	0.00
Nepal	-3.0341	0.0000	0.0662	0.00
India	-2.0244	0.0000	0.0496	0.00
South Africa	-5.1070	0.0000	0.0709	0.00
Brazil	-4.5529	0.0000	-0.0324	0.00
UK	-4.5630	0.0020	0.0031	0.00
Japan	-5.9846	0.0000	-0.0422	0.00
USA	-3.7902	0.0000	0.0295	0.00
Germany	-5.6408	0.0000	-0.0074	0.00
	N =324	R^2= 0.46	N =312	R^2= 0.9857

8.4.5 Panel Unit root test

Increasingly recent studies have looked into nonstationarity and heterogeneity issues in panel data model. Levin and Lin (1992)

$$\Delta y_{i,t} = \alpha_i + \rho y_{i,t-1} + \sum_{k=1}^{n} \phi_k \Delta y_{i,t-1} + \delta_i t + \theta_t + u_{i,t} \tag{1450}$$

$H_0 : \rho = 1$ against $H_0 : \rho < 1$

Levin, A., C. Lin and C. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and finite sample properties", Journal of Econometrics, 108, p.12-24.

IM, Pesharan and Shin (1997)

Im, K.S., M. Pesaran and Y. Shin (2003): "Testing for Unit Roots in Heterogeneous Panels", Journal of Econometrics, 115, p.53-74.

$$\Delta y_{i,t} = \alpha_i + \rho y_{i,t-1} + \sum_{k=1}^{n} \phi_i \Delta y_{i,t-1} + \delta_i t + \theta_t + u_{i,t} \tag{1451}$$

Heterogeneity in unit roots: against no unit root

$$t_{IPS} = \frac{\sqrt{N}\left(t - \frac{1}{N}\sum\limits_{k=1}^{n} E\left[t_{iT}|\rho_i = 0\right]\right)}{\sqrt{\frac{1}{N}\sum\limits_{k=1}^{n} var\left[t_{iT}|\rho_i = 0\right]}} \Longrightarrow N(0,1) \tag{1452}$$

- Panel Unit root test

Table 22: Cointegration Test of Growth and Exchange Rate Equations

	Cointegration in Growth Model		Cointegration in Exchange rate Model	
ADF test (T=321; Constant; 5%=-2.87; 1% = -3.45)				
Determinants	ADF-Statistics	Decision	ADF-Statistics	Decision
Investment ratio	-4.449**	Stationary	-	-
Export Ratio	-1.9000	Non-Stationary	-	-
Exchange rate -1	-	-	-1.510	Non-Stationary
Real Interest rate	-	-	-2.59	Non-Stationary
Population growth rate	-6.171**	Stationary	-6.171**	Stationary
Residual	-10.62**	Stationary	-4.96**	Stationary
Conclusion: Variables in both growth and exchange rates equations are cointegrated.				

Maddala and Wu (1999) tests for unbalanced panel

Maddala, G.S. and S. Wu (1999): "A comparative study of unit root test with panel data and a new simple test", Oxford Bulletin of Economics and Statistics, 61, p.631-652.

$\Pi = -2\sum_{k=1}^{n} \ln \pi_i$ with χ^2 distribution

where π_i is the probability limit of ADF test

KSSS test

$$KPSS = \sum_{t=1}^{T} \frac{S_t^2}{\widehat{\sigma}^2} \tag{1453}$$

where $S_t^2 = \sum_{t=1}^{T} e_t$ is the partial sum of errors in a regression of Y on an intercept and time trend. In contrast to the unit root test this test assumes that Y series are stationary and alternative hypothesis is nonstationarity.

Kwaitkowski, D. P.C. Phillips, P. Schmidt and Y. Shin (1992): "Testing the null hypothesis of Stationarity against the alternative of a unit root", Journal of Econometrics, 54, p.159-178.

- Panel Unit root: Kao test

Kao, C. (1999): "Spurious Regression and residual-based Tests for Cointegration in Panel Data", Journal of Econometrics, 90, p.1-44.

Start with $y_{i,t} = \alpha_i + x_{i,t}\beta + e_{i,t}$ $\qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right)$

Residual based cointegration $e_{i,t} = \rho e_{i,t-1} + v_{i,t}$

Estimate $\rho = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N} \widehat{e}_{i,t}\widehat{e}_{i,t-1}}{\sum_{t=1}^{T}\sum_{n=1}^{N} \widehat{e}_{i,t}^2}$ and

related t statistics $t_p = \frac{(\widehat{\rho}-1)\sqrt{\sum_{t=1}^{T}\sum_{n=1}^{N} \widehat{e}_{i,t}^2}}{\frac{1}{NT}\left(\sum_{t=1}^{T}\sum_{n=1}^{N}\left(\widehat{e}_{i,t}^2 - \widehat{\rho}\widehat{e}_{i,t}^2\right)\right)}$

Panel Unit Root Test in Eviews

Panel unit root test: Summary
Series: EDU_R
Date: 04/22/10 Time: 08:55
Sample: 1971 2006
Exogenous variables: Individual effects
User-specified lags: 1
Newey-West automatic bandwidth selection and Bartlett kernel
Balanced observations for each test

Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-2.57787	0.0050	14	476
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-0.17536	0.4304	14	476
ADF - Fisher Chi-square	23.2299	0.7215	14	476
PP - Fisher Chi-square	28.7105	0.4273	14	490

Panel Cointegration in Eviews

Save data in excel/csv; import in Eviews as foreign data file/ Select Basic structure as panel data (have panel id and year id variables in the data file); Quick/ Group statistics/Johansen cointegration test; then list variables; select pedroni (Engle-Granger based) – select other specifications then estimate. You get results as following.

Study the trace and max –eigen value tests.

Unrestricted Cointegration Rank Test (Trace and Maximum Eigenvalue)

Hypothesized No. of CE(s)	Fisher Stat.* (from trace test)	Prob.	Fisher Stat.* (from max-eigen test)	Prob.
None	72.00	0.0000	65.75	0.0001
At most 1	27.83	0.4735	22.71	0.7477
At most 2	20.99	0.8256	17.79	0.9314
At most 3	32.79	0.2435	32.79	0.2435

8.5 Panel Cointegration

Analytical results from a dynamic optimisation model for a global economy show how exchange rates are determined by relative prices of trading countries.

Prices depend on preferences on domestic and foreign goods, marginal productivities of capital and labour as well as the relative rates of taxes and tariffs across two countries.

Dynamic model is solved for numerical simulation and scenario analyses. Long run relationship obtained in the dynamic general equilibrium are tested by the GMM estimation of dynamic panel model.

The determinants of growth of per capita output and the exchange rates across eleven countries representing the global economy in fact validate the conclusion of general equilibrium results.

Estimates support the standard neoclassical theory of economic growth and uncovered interest parity theory of exchange rate though country specific factors, including preferences and technology, can also have significant influence in estimation of each of these models.

- Panel Cointegration: Larsson Test:

Based their test on Johanhes' maxmimum likelihood procedure.

$\Delta Y_{i,t} = \Pi_i Y_{i,t-1} + \sum\limits_{k=1}^{n} \Gamma_k \Delta y_{i,t-k} + u_{i,t}$

$H_0 : rank\left(\Pi_i\right) - r_i < r$ for all i from 1 to N.

$H_A : rank\left(\Pi_i\right) = p$ for all i from 1 to N.

The standard rank test statistics is defined in terms of average of the trace statistic for each cross section unit and mean and variance of trace statistics.

$$LR = \frac{\sqrt{N}\left(LR_{NT} - E\left(Z_k\right)\right)}{\sqrt{var\left(Z_t\right)}} \tag{1454}$$

- Panel Cointegration: PedroniTest
- Within group tests: Panel v statistic

$$T^2 N^3 Z_{vNT} = \frac{T^2 N^{\frac{3}{2}}}{\sum\limits_{t=1}^{T}\sum\limits_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t}^2\right)} \tag{1455}$$

Panel ρ statistic

$$T\sqrt{N} Z_{\rho NT} = \frac{T\sqrt{N}\left(\sum\limits_{t=1}^{T}\sum\limits_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t}^2\right)\Delta\widehat{e}_{i,t}^2 - \widehat{\lambda}_i\right)}{\sum\limits_{t=1}^{T}\sum\limits_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t}^2\right)} \tag{1456}$$

Panel t statistic

$$Z_{tNT} = \sqrt{\sigma_{NT}^2 \sum_{t=1}^{T}\sum_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t-1}^2\right)}\left(\sum_{t=1}^{T}\sum_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t}^2\right)\Delta\widehat{e}_{i,t}^2 - \widehat{\lambda}_i\right) \tag{1457}$$

Panel t statistic (parametric)

$$Z_{tNT} = \sqrt{\sigma_{NT}^2 \sum_{t=1}^{T}\sum_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t-1}^2\right)}\left(\sum_{t=1}^{T}\sum_{n=1}^{N} L_{1,1}^{-2}\left(\widehat{e}_{i,t}^2\right)\Delta\widehat{e}_{i,t}^2 - \widehat{\lambda}_i\right) \tag{1458}$$

- Between group tests

Group statistic

$$T\sqrt{N}Z_{\rho NT} = \frac{T\sqrt{N}\sum_{t=1}^{T}\left(\widehat{e}_{i,t}^{2}\Delta\widehat{e}_{i,t}^{2} - \widehat{\lambda}_i\right)}{\sum_{t=1}^{T}\sum_{n=1}^{N}\left(\widehat{e}_{i,t}^{2}\right)} \tag{1459}$$

Group t statistic

$$\sqrt{N}Z_{tNT-1} = \sqrt{N}\sum_{n=1}^{N}\sqrt{\sigma_i^2\sum_{t=1}^{T}\widehat{e}_{i,t}^{2}}\sum_{t=1}^{T}\left(\widehat{e}_{i,t}^{2}\Delta\widehat{e}_{i,t}^{2} - \widehat{\lambda}_i\right) \tag{1460}$$

Group t statistic (parametric)

$$\sqrt{N}Z_{tNT-1} = \sqrt{N}\sum_{n=1}^{N}\sqrt{\sigma_i^2\sum_{t=1}^{T}\widehat{e}_{t}^{2}}\sum_{t=1}^{T}\left(\widehat{e}_{i,t}^{2}\Delta\widehat{e}_{i,t}^{2} - \widehat{\lambda}_i\right) \tag{1461}$$

8.5.1 Panel Model for Limited Dependent Variables

Panel models of limited dependent variables
$y_{i,t}^{*} = \alpha_i + x_{i,t}\beta + e_{i,t}$ $\qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right)$
$y_{i,t} = 1$ if $y_{i,t}^{*} > 0$ where $y_{i,t}^{*}$is a latent variable; $y_{i,t} = 0$ otherwise.

$$\begin{aligned}\log L\left(\beta, \alpha_1, ..., \alpha_N\right) &= \sum_{i,t} y_{i,t}\log F\left(\alpha_i + x_{i,t}'\beta\right) \\ &\quad + \sum_{i,t}\left(1 - y_{i,t}\right)\left(1 - \log F\left(\alpha_i + x_{i,t}'\beta\right)\right)\end{aligned} \tag{1462}$$

$$\begin{aligned}&f\left(y_{i,t}, ..., y_{i,T}/x_{i,t}, ..., x_{i,T}, \beta\right) \\ = &\int_{-\infty}^{\infty} f\left(y_{i,t}, ..., y_{i,T}/x_{i,t}, ..., x_{i,T}, \beta\right) f\left(\alpha_i\right) d\alpha_i \\ = &\int_{-\infty}^{\infty}\left[\prod_i f\left(y_{i,t}/x_{i,t}, \beta\right)\right] f\left(\alpha_i\right) d\alpha_i\end{aligned} \tag{1463}$$

Random Effect Tobit Model
$y_{i,t}^{*} = \alpha_i + x_{i,t}\beta + e_{i,t}$ $\qquad e_{i,t} \sim IID\left(0, \sigma_e^2\right)$
$y_{i,t} = 1$ if $y_{i,t}^{*} > 0$ where $y_{i,t}^{*}$is a latent variable; $y_{i,t} = 0$ otherwise.

$$\begin{aligned}&f\left(y_{i,t}, ..., y_{i,T}/x_{i,t}, ..., x_{i,T}, \beta\right) \\ = &\int_{-\infty}^{\infty} f\left(y_{i,t}, ..., y_{i,T}/x_{i,t}, ..., x_{i,T}, \beta\right) f\left(\alpha_i\right) d\alpha_i \\ = &\int_{-\infty}^{\infty}\left[\prod_i f\left(y_{i,t}/x_{i,t}, \beta\right)\right] f\left(\alpha_i\right) d\alpha_i\end{aligned} \tag{1464}$$

$$f\left(y_{i,t}/x_{i,t},\beta\right) = \Phi\left(\frac{x'_{i,t}\beta+\alpha_i}{\sqrt{1-\sigma_u^2}}\right) \quad \text{if} \quad y_{i,t}=1 \tag{1465}$$

$$.......... = (1-\Phi)\left(\frac{x'_{i,t}\beta+\alpha_i}{\sqrt{1-\sigma_u^2}}\right) \quad \text{if} \quad y_{i,t}=0 \tag{1466}$$

Dynamic Tobit Panel Model

$$y^*_{i,t} = \alpha_i + \gamma y_{i,t-1} + x'_{i,t}\beta + e_{i,t} \tag{1467}$$

$$\begin{aligned} & f\left(y_{i,t},.....,y_{i,T}/x_{i,t},.....,x_{i,T},\beta\right) \\ = & \int_{-\infty}^{\infty} f\left(y_{i,t},.....,y_{i,T}/x_{i,t},.....,x_{i,T},\beta\right) f\left(\alpha_i\right) d\alpha_i \\ = & \int_{-\infty}^{\infty}\left[\prod_i f\left(y_{i,t}/y_{i,t-1},x_{i,t},\alpha_i,\beta\right)\right] \\ & f\left(y_{i,t}/x_{i,t},\beta\right) f\left(\alpha_i\right) d\alpha_i \end{aligned} \tag{1468}$$

$$\begin{aligned} f\left(y_{i,t}/y_{i,t-1},x_{i,t},\beta\right) &= \Phi\left(\frac{x'_{i,t}\beta+\gamma y_{i,t-1}+\alpha_i}{\sqrt{1-\sigma_u^2}}\right) \quad \text{if} \quad y_{i,t}=1 \\ .... &= (1-\Phi)\left(\frac{x'_{i,t}\beta++\gamma y_{i,t-1}+\alpha_i}{\sqrt{1-\sigma_u^2}}\right) \\ \text{if} \quad y_{i,t} &= 0 \end{aligned} \tag{1469}$$

8.5.2 Error Component Method

The error component method decomposes these errors into a common intercept and the random part.

Thus the model will take the following form:

$$y_{i,t} = \alpha_{0,i} + \alpha_{1,i}x_{i,t} + e_{i,t} \tag{1470}$$

$$\alpha_{0,i} = \overline{\alpha}_1 + \mu_i \tag{1471}$$

where i =1.., N. Each cross section unit (country) had its own intercept parameter in the pooled dummy variable model.

$\overline{\alpha}_1$ represents the population mean intercept and μ_i are independent of each error $e_{i,t}$. It has a constant mean and constant variance. $E\left(\mu_i\right)=0; var\left(\mu_i\right)=\sigma_\mu^2$

$$y_{i,t} = \left(\overline{\alpha}_1+\mu_i\right)+\alpha_{1,i}x_{i,t}+e_{i,t} \tag{1472}$$

$$= \overline{\alpha}_1+\alpha_{1,i}x_{i,t}+\left(e_{i,t}+\mu_i\right) = \overline{\alpha}_1+\alpha_{1,i}x_{i,t}+v_{i,t} \tag{1473}$$

$$v_{i,t} = e_{i,t} + \mu_i \tag{1474}$$

8.5.3 Error Component Method

The error component include overall error and individual specific error , : common and individual specific errors.

the compound error term has mean zero, $E\left(v_{i,t}\right) = 0$;

its variance $var\left(v_{i,t}\right) = \sigma_{\mu}^2 + \sigma_e^2$ is homoskedastic

$covar\left(v_{i,t}, v_{i,s}\right) = \sigma_{\mu}^2$ error from the same country in different periods are correlated

$covar\left(v_{i,t}, v_{j,s}\right) = \sigma_{\mu}^2$ for i=j errors from different countries are always uncorrelated.

Like in the SURE method the generalised least square estimator, with transformed method produces the most efficient estimators or error component model.

The error component include overall error and individual specific error , : common and individual specific errors.

the compound error term has mean zero, $E\left(v_{i,t}\right) = 0$;

its variance $var\left(v_{i,t}\right) = \sigma_{\mu}^2 + \sigma_e^2$ is homoskedastic

$covar\left(v_{i,t}, v_{i,s}\right) = \sigma_{\mu}^2$ error from the same country in different periods are correlated

$covar\left(v_{i,t}, v_{j,s}\right) = \sigma_{\mu}^2$ for $i = j$ errors from different countries are always uncorrelated.

Like in the SURE method the generalised least square estimator, with transformed method produces the most efficient estimators or error component model.

References

[1] Arellano, M., and S. Bond (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies* 58, 277–297.

[2] Arellano M and S. Bonhomme (2012) Identifying Distributional Characteristics in Random Coefficients Panel Data Models, Review of Economic Studies 79 (3): 987-1020

[3] Arellano, M., and O. Bover (1995) 'Another look at the instrumental variable estimation of error-components models', *Journal of Econometrics* 68, 29–51.

[4] Arulampalam Wiji, Alison L. Booth and Mark L. Bryan (2004) Training and the New Minimum Wage, The Economic Journal, 114, 494, C87-C94

[5] Arulampalam W., Booth A.L., (1998), "Learning and Earning: Do Multiple Training Events Pay?: A Decade of Evidence from a Cohortof Young British Men", Economica, 68, 271, 379-400

[6] Arulampalam Wiji and Mark B. Stewart (1995) The Determinants of Individual Unemployment Durations in an Era of High Unemployment The Economic Journal, 105, 429, 321-332

[7] Baltagi B.H, Qu Feng, Chihwa Kao (2012) A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model Journal of Econometrics, 170, 1, 164-177

[8] Baltagi H. (2008) Econometric analysis of panel data, Fourth Ediction, Blackwell.

[9] Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. Journal of Business and Economic Statistics 23, 49–60

[10] Basu P, K. Bhattarai and Y. Gatechu (2019) Capital adjustment cost and bias in income based dynamic panel models with fixed effects, German Economic Review,20(4), pp.e1002-e1018.

[11] Basu P. and K. Bhattarai (2012) Cognitive Skills, Openness and Growth, Economic Record, 88: 280: March, 18-38

[12] Bhargava A (1991) Identification and Panel Data Models with Endogenous Regressors Review of Economic Studies 58 (1): 129-140

[13] Blundell, R. and S. Bond (1998) 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics,* 87, 115–143.

[14] Blundell R W and R J. Smith (1989) Estimation in a class of simultaneous equation limited dependent variable models, Review of Economic Studies, 56:37-38.

[15] Caner M and B. E. Hansen Instrumental Variable Estimation Of A Threshold Model, Econometric Theory, 20, 2004, 813–843

[16] Carriero A, Kapetanios G and Marcellino M. 2009. Forecasting exchange rates with a large Bayesian VAR.International Journal of Forecasting 25: 400-417

[17] Cornwell C and P. Rupert (1988) Efficient Estimation With Panel Data: An Empirical Comparison of Instrumental Variables Estimators Journal of Applied Econometrics, 3, 2, 149-155

[18] Chesher A (1984) Improving the efficiency of Probit estimators, Review of Economic Studies,66:3:523-527.

[19] Hall A (2003) Generalised Methods of Moments in Baltagi ed. Theoretical Econometrics, Blackwell.

[20] Hansen, B.E. (1999) Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference, Journal of Econometrics, 93, 345-68.

[21] Hansen L.P. (1982) Large sample properties of generalized method of moment estimators, Econometrica, 50:4:1029-1054.

[22] Hsiao C. (1993) Analysis of panel data, Cambridge.

[23] Hausman J.A., (1978), "Specification Tests in Econometrics", Econometrica, Vol. 46, No. 6, pp.1251-1271.

[24] Hayakawa K (2009) On the effect of mean-nonstationarity in dynamic panel data models, Journal of Econometrics, 153, 2, 133-135

[25] Heckman J. J., (1979), Sample Selection Bias as a Specification Error, Econometrica, Vol. 47, No. 1, pp153-161.

[26] Im, K.S., M. Pesaran and Y. Shin (2003): "Testing for Unit Roots in Heterogeneous Panels", Journal of Econometrics, 115, p.53-74.

[27] Imbens G. W. and T Lancaster (1994) Combining Micro and Macro Data in Microeconometric Models, Review of Economic Studies, 61:4:655-680.

[28] Johnston J. (1960) Econometric methods, 7th edition, McGraw Hill.

[29] Johansen S. (1988) Estimation and Hypothesis Testing of Cointegration Verctors in Gaussian Vector Autoregressive Models, Econometrica, 59:6, 1551-1580.

[30] Kapetanios G, J. Mitchell and Y. Shin (2014) A nonlinear panel data model of cross-sectional dependence Journal of Econometrics, Available online

[31] Keifer N (1988) Economic duration data and hazard functions, Journal of Economic Literature, 26:647-679.

[32] Kao, C. (1999): Spurious Regression and residual-based Tests for Cointegration in Panel Data, Journal of Econometrics, 90, p.1-44.

[33] Kwaitkowski, D. P.C. Phillips, P. Schmidt and Y. Shin (1992): Testing the null hypothesis of Stationarity against the alternative of a unit root, Journal of Econometrics, 54, p.159-178.

[34] Kyriazidou E (2001) Estimation of Dynamic Panel Data Sample Selection Models,Review of Economic Studies 68 (3): 543-572

[35] Kleibergen, F. and Paap, R., (2006). Generalized reduced rank tests using the singular value decomposition, Journal of Econometrics 133, 97-126

[36] Koop G.M. (2013) Forecasting With Medium And Large Bayesian VARS, Journal of Applied Econometrics 28, 2, 177-203

[37] Lancaster T (2002) Orthogonal Parameters and Panel Data Review of Economic Studies 69 (3): 647-666

[38] Lancaster T and A Chesher (1983) The Estimation of Models of Labour Market Behviour Review of Economic Studies, 50:4:609-624.

[39] Larsson, R., J. Lyhagen and M. Lothgren (2001): Likelihood-based cointegration tests in heterogeneous panels, Econometrics Journal, 4, p.109-142

[40] Levin, A., C. Lin and C. Chu (2002): "Unit Root Tests in Panel Data: Asymptotic and finite sample properties", Journal of Econometrics, 108, p.12-24.

[41] Lee Y (2014) Testing a linear dynamic panel data model against nonlinear alternatives, Journal of Econometrics, 178, Part 1, 146-166

[42] McCoskey, S. and C. Kao (1999): "Testing the stability of a Production function with Urbanisation as a shift factor", Oxford Bulletin of Economics and Statistics, 61, p.671-690.

[43] Maddala, G.S. and S. Wu (1999): A comparative study of unit root test with panel data and a new simple test, Oxford Bulletin of Economics and Statistics, 61, p.631-652.

[44] Pedroni, P. (1999): "Critical values for cointegration tests in heterogeneous panels with multiple regressors", Oxford Bulletin of Economics and Statistics, 61, p.653-670.

[45] Pesaran, M.H. and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels", Journal of Econometrics, 68, p.79-113

[46] Phillips P.C.B. (1987) Time Series Regression with an Unit Root, Econometrica, vol. 55, No. 2, 277-301.

[47] Phillips P.C.B. and D Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence, Journal of Econometrics 137, 162–188

[48] Robertson D. and J. Symons (1992) Some Strange Properties of Panel Data Estimators, Journal of Applied Econometrics, 7, 2, 175-189

[49] Sentana, E. (2009), The econometrics of mean-variance efficiency tests: a survey. The Econometrics Journal, 12: C65–C101

[50] Semykina A. and J M. Wooldridge (2013) Estimation Of Dynamic Panel Data Models With Sample Selection, Journal of Applied Econometrics, 28, 1 , 47-61

[51] Staigler D. and Stock J. H., (1997), "Instrumental Variables Regression with Weak Instruments", Econometrica, 65, 3, pp.557-586.

[52] Su L and X Lu (2013) Nonparametric dynamic panel data models: Kernel estimation and specification testing Journal of Econometrics, 176, 2, 112-133

[53] Tsionas, M. ed., 2019. Panel Data Econometrics: Empirical Applications. Academic Press.

[54] Windmeijer, F (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators, Journal of Econometrics 126: 25–51.

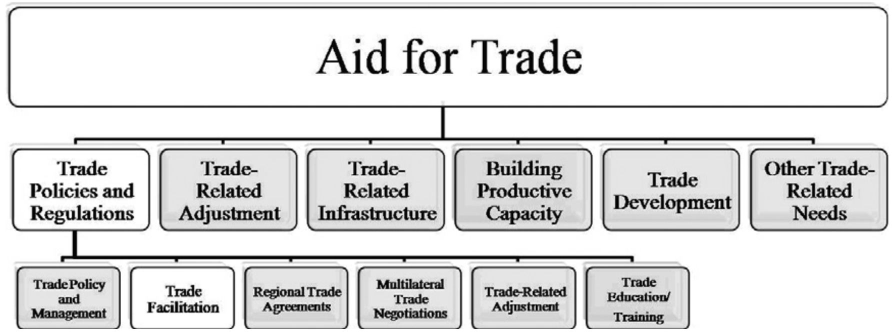
[55] Wooldridge J.M (2005) Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity, Journal of Applied Econometrics, 20, 1, 39-54

[56] Wooldridge J. M. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press.

8.6 Panel Data Method for Analysing Link Between Trade and Aid

Literature on aid and trade could be updated

- - Bandyopadhyay, S. and Vermann, E. (2012), Cali, M., Razzaque, M. and te Velde, D.W. (2011),
  - Deardorff, A. and Stern, R. (2009), Gounder, R. (1995), Gounder, R. (1994), Helble, M., Mann, C., Wilson, J. (2012) , Lundsgaarde, E., Breunig, C. and Prakash, A. (2010) , Morrissey, O. (1993) , Nitsch, V. (2000) Stiglitz, J. and Charlton, A. (2013) Stiglitz, J. and Charlton, A. (2006) , Vijil, M. and Wagner, D. (2012), Wagner, D. (2003)
- new literature
  - Alain de Janvry and Elisabeth Sadoulet (1988), Georgios Karras (2004) , Tanweer Akram (2003)
  - Nils B. Weidmann , Doreen Kuse & Kristian Skrede Gleditsch (2010), Chun-Chieh Wang (2011)



- Matthias Busse, Ruth Hoekstra and Jens Königer (2012), Takumi Naito (2012), Vijil M and Laurent Wagner (2012) Rifat Baris Tekin (2012) , Juhasz Silva and Douglas Nelson (2012), Jean-Jacques Hallaert (2013) , Jan Pettersson and Lars Johansson (2013), Gamberoni and Richard Newfarmer (2014) , Olivier Cadot, Ana Fernandes, Julien Gourdon, Aaditya Mattoo and Jaime de Melo (2014), Mariana Vijil (2014)

Wagner equation

$$\begin{aligned}\ln(T_{dr}) &= \beta_1 \ln\left(\frac{Y_d Y_r}{Y_W}\right) + \beta_2 \ln\left(\frac{Y_d}{P_d}\right) + \beta_3 \ln\left(\frac{Y_r}{P_r}\right) + \beta_4 \ln(D_{dr}) \\ &\quad + \beta_5 \ln(REM_d) + \beta_6 \ln(REM_r) + \beta_7 \ln(LAN_{dr}) + \varepsilon_{dr}\end{aligned}$$

$$\begin{aligned}\ln(A_{dr}) &= \beta_1 \ln(Y_d) + \beta_1 \ln(Y_r) + \beta_3 \ln(D_{dr}) + \beta_4 \ln\left(\frac{Y_d}{P_d}\right) \\ &\quad + \beta_5 \ln\left(\frac{Y_r}{P_r}\right) + \beta_6 \ln(LAN_{dr}) + \beta_7 \ln(MILSR_{dr}) + C + \varepsilon_{dr}\end{aligned}$$

$$\ln(T_{dr}) = \ln\Gamma_{dr} + \beta_8 \ln(\max\{1, A_{dr}\}) + \beta_9 \ln(NAD_{dr}) + \varepsilon_{dr}$$

$REM_r = \frac{1}{\sum_d \left(\frac{\left(\frac{Y_d}{Y_W}\right)}{D_{dr}}\right)}$

http://www.distancefromto.net/countries.php; http://www.cepii.fr/CEPII/en/bdd_modele/models.asp

Matthias Busse, Ruth Hoekstra and Jens Königer(2012) Kristian Skrede Gleditsch (2002) and Mayer, T. and Zignago, S. (2006) the great circle distance between capital cities.

Trade Impacts of Aid for the UK: Do file and panel settings

- import excel "C:\AIEFS\tdata\Aidpanel_UK_2014.xlsx", sheet("Sheet1") firstrow
- xtset ID tt, yearly
- generate aidd1 = Aid+1
- xtreg exp YYUKYA ypuk dist Aid, re

```
. xtdpdsys exp YYUKYA ypuk yp dist aidd1, lags(1) twostep artests(2)
note: dist dropped from div() because of collinearity

System dynamic panel-data estimation        Number of obs        =       216
Group variable: ID                          Number of groups     =        22
Time variable: tt
                                            Obs per group:    min =         2
                                                              avg =  9.818182
                                                              max =        11

Number of instruments =     70              Wald chi2(3)         =  6.55e+08
                                            Prob > chi2          =    0.0000
Two-step results
------------------------------------------------------------------------------
         exp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |
         L1. |   1.005088   .0002277  4413.13   0.000     1.004642    1.005535
             |
      YYUKYA |  -1.79e-06   1.00e-08  -178.52   0.000    -1.81e-06   -1.77e-06
        ypuk |   3.689713   .0573206    64.37   0.000     3.577367     3.80206
          yp |          0  (omitted)
        dist |          0  (omitted)
       aidd1 |   1.985273   .0379436    52.32   0.000     1.910905    2.059642
       _cons |          0  (omitted)
------------------------------------------------------------------------------
Warning: gmm two-step standard errors are biased; robust standard
         errors are recommended.
Instruments for differenced equation
        GMM-type: L(2/.).exp
        Standard: D.YYUKYA D.ypuk D.yp D.aidd1
Instruments for level equation
        GMM-type: LD.exp
        Standard: _cons
```

- xtreg exp YYUKYA ypuk dist Aid, fe
- xtdpdsys exp YYUKYA ypuk yp dist aidd1, lags(1) twostep artests(2)
- xtreg lexp lyyuka lyp lypuk ldist laidd1, re
- xtreg lexp lyyuka lyp lypuk ldist laidd1, fe
- xtabond lexp lyyuka lyp lypuk ldist laidd1, lags(1) twostep artests(2)
- Data sources: exp YYUKYA ypuk yp WBDI; distance from the Google map and Kristian Skrede Gleditsch (2002); AID from OECD.

Trade Impacts of Aid for the UK: Random effect Panel Estimation
Trade Impacts of Aid for the UK: Fixed effect Panel Estimation
Trade Impacts of Aid for the UK: Dynamic Panel Model Two Stage Estimation

8.6.1 Estimations in log forms for elasticities

Trade Elasticity Impacts of Aid for the UK: Random effect Panel Estimation
Trade Elasticity Impacts of Aid for the UK: Fixed effect Panel Estimation
Hausemann test supports random effect model
Trade Elasticity Impacts of Aid for the UK: Dynamic Panel Model Two Stage Estimation

References

[1] Akram Tanweer (2003) The international foreign aid regime: who gets foreign aid and how much?Applied Economics, 35:11, 1351-1356

[2] Bandyopadhyay, S. and Vermann, E. (2012) “Donor Motives for Foreign Aid”, Federal Reserve Bank of St. Louis Review, Vol. 95, No. 4, pp. 327-36.

```
. xtabond lexp lyyuka lyp lypuk ldist laidd1, lags(1) twostep artests(2)
note: ldist dropped from div() because of collinearity

Arellano-Bond dynamic panel-data estimation  Number of obs         =       133
Group variable: ID                           Number of groups      =        21
Time variable: tt
                                             Obs per group:    min =         1
                                                               avg =  6.333333
                                                               max =        10

Number of instruments =     60               Wald chi2(6)          = 155168.16
                                             Prob > chi2           =    0.0000
Two-step results
------------------------------------------------------------------------------
        lexp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lexp |
         L1. |  -.1796201   .0388037    -4.63   0.000     -.255674   -.1035662
             |
      lyyuka |  -1.562127   .8335513    -1.87   0.061    -3.195857    .0716038
         lyp |   2.764037    .346676     7.97   0.000     2.084564    3.443509
       lypuk |   1.766329   .6814275     2.59   0.010     .4307552    3.101902
       ldist |   1.309397   1.255414     1.04   0.297    -1.151169    3.769964
      laidd1 |  -.0252523   .0033013    -7.65   0.000    -.0317226   -.0187819
       _cons |          0  (omitted)
------------------------------------------------------------------------------
Warning: gmm two-step standard errors are biased; robust standard
         errors are recommended.
Instruments for differenced equation
        GMM-type: L(2/.).lexp
        Standard: D.lyyuka D.lyp D.lypuk D.laidd1
Instruments for level equation
        Standard: _cons
```


[3] Basnett, Y., Engel, J., Kennan, J., Kingombe, C., Massa, I., te Velde, D. (2012) "Increasing the Effectiveness of Aid for Trade: The Circumstances Under Which it Works Best", ODI Working Paper 353, Overseas Development Institute, London.

[4] Busse, M., Hoekstra R. and Königer (2011) The Impact of Aid for Trade Facilitation on the Costs of Trading, Ruhr-University of Bochum, Germany, mimeo.

[5] Busse Matthias, Ruth Hoekstra and Jens Königer (2012) The Impact of Aid for Trade Facilitation on the Costs of Trading, KYKLOS, Vol. 65 – May 2012 – No. 2, 143–163

[6] Cadot Olivier , Ana Fernandes, Julien Gourdon, Aaditya Mattoo and Jaime de Melo (2014) Evaluating Aid for Trade: A Survey of Recent Studies, The World Economy (2014)

[7] Cali, M., Razzaque, M. and te Velde, D.W. (2011) Effectiveness of Aid for Trade in Small and Vulnerable Economics, London: Commonwealth Secretariat.

[8] Cali, M. and te Velde, D.W. (2009) Does aid for trade really improve trade performance? Overseas Development Institute, London.

[9] Calì, M. and te Velde, D.W. (2008) "The Effectiveness of Aid for Trade: Some Empirical Evidence", Trade Hot Topics 50, Commonwealth Secretariat, London.

[10] Commonwealth Government (2012) The Association of Southeast Asian Nations (ASEAN)-Australia-New Zealand Free Trade Area (AANZFTA), Canberra: AGPS.

[11] de Janvry Alain and Elisabeth Sadoulet (1988) The Conditions for Compatibility between Aid and Trade in Agriculture, Economic Development and Cultural Change, Vol. 37, No. 1 (Oct., 1988), pp. 1-30

[12] Deardorff, A. and Stern, R. (2009) Alternatives to the Doha Round. Journal of Policy Modeling, 31(4), pp. 526-39.

[13] Gamberoni and Richard Newfarmer (2014) Aid for Trade: Do Those Countries that Need it, Get it? World Economy (2014) 542-554

[14] Gounder, R. (1995) "Non-Nested Models of Australia's Overseas Aid Program", Applied Economics, Vol. 27(7), 609-21.

[15] Gounder, R. (1994) "Empirical Results of Aid Motivations: Australia's Bilateral Aid Programs", World Development, Vol. 22(1), 99-113.

[16] Hallaert Jean-Jacques (2013) The future of Aid for Trade: challenges and options World Trade Review (2013), 12: 4, 653–668

[17] Head, K., Ries, J. and Wagner, D. (1998) "Immigrants and the Trade of Provinces", Research on. Immigration and Integration in the Metropolis (RIIM), Discussion Paper, No. 98-21, Vancouver, Canada.

[18] Helble, M., Mann, C., Wilson, J. (2012) "Aid for Trade Facilitation", Review of World Economics, Vol. 148(2), 357-76.

[19] Hoekman, B. (2010) Aid for Trade: Why, What and Where are We? Washington, DC, World Bank. Retrieved from

[20] Hoekman, B. and Prowse, S. (2009). Economic Policy Responses to Preference Erosion: From Trade as Aid to Aid for Trade. In B. Hoekman, W. Martin and P. Braga (Eds.), Trade Preference Erosion-Measurement and Policy Response.

[21] Washington, DC, Palgrave Macmillan and World Bank, Chapter 11, pp. 428-448.

[22] Karras Georgios (2004) Aid versus Trade: Comparing their Growth Effects for a Panel of Developing Countries,The Journal Of Economic Asymmetries December 2004

[23] Leamer, E. (1977) Access to Western Markets and Eastern Effort Levels, in Zecchini, S. (ed.) Lessons from the Economics Transition: Central and Eastern Europe in the 1990s, Boston: Kluwer Academic.

[24] Lundsgaarde, E., Breunig, C. and Prakash, A. (2010) "Instrumental Philanthropy: Trade and the Allocation of Foreign aid". Canadian Journal of Political Science Vo;.43(3), 733-761.

[25] Mayer, T. and Zignago, S. (2006) CEPII's Distance Measures, Centre for Economic Policy Research (CEPR), Paris.

[26] Morrissey, O. (1993) "The Mixing of Aid and Trade Policies", The World Economy, Vol. 15(1), 69-84.

[27] Nilsson, L. (1998) Aid and Donor Exports: The Case of the EU Countries, mimeo, Lund.

[28] Nitsch, V. (2000) "National Boarders and International Trade: Evidence from the European Union", Canadian Journal of Economics, Vol. 22(4), 1019-35.

[29] Organisation for Economic Cooperation and Development (2015) OECD Library Database, Creditor Reporting System

[30] Pettersson Jan and Lars Johansson (2013) Aid, Aid for Trade, and bilateral trade: An empirical study,The Journal of International Trade & Economic Development, 2013 Vol. 22, No. 6, 866–894

[31] Silva Simone Juhasz and Douglas Nelson (2012)Does Aid Cause Trade? Evidence from an Asymmetric Gravity Model, The World Economy (2012), 545-576

[32] Stiglitz, J. and Charlton, A. (2013) The Right to Trade, Rethinking the Aid for Trade Agenda, London: Commonwealth Secretariat.

[33] Stiglitz, J. and Charlton, A. (2006) "Aid for Trade", International Journal of Development Issues Vol. 5(2), 1-41.

[34] Mariana Vijil (2014) Aid for Trade Effectiveness: Complementarities with Economic Integration, The World Economy (2014), 555-566

[35] Vijil, M. and L Wagner (2012) "Does Aid for Trade Enhance Export Performance? Investigating the Infrastructure Channel", The World Economy, Vol. 35(7), 838-68.

[36] Takumi Naito (2012) Aid For Trade And Welfare, The Japanese Economic Review Vol. 63, No. 2, June 2012

[37] Tekin Rifat Baris (2012) Development aid, openness to trade and economic growth in Least Developed Countries: bootstrap panel Granger causality analysis, Social and Behavioral Sciences 62 (2012) 716 – 721

[38] Wagner, D. (2003) "Aid and Trade – An Empirical Study", Journal of the Japanese and International Economies, Vol. 17, 153-73.

[39] World Bank (2015) World Development Indicators, www.worldbank.org/wdi/data, Washington D.C.: World Bank.

[40] Wang Chun-Chieh (2011) Aid for trade as a public good, The Journal of International Trade & Economic Development, Vol. 20, No. 6, December 2011, 711–728

[41] Weidmann Nils B., Doreen Kuse & Kristian Skrede Gleditsch (2010) The Geography of the International System: The CShapes Dataset, International Interactions, 36:1,86-106.

9 L8: Duration Analysis

Many economic events such as unemployment or working life or price rises. Durnations models have been applied to analyse such events by Lancaster (1979), Elbers and Ridder (1982), Heckman and Singer (1984), Kennan (1985), Struthers and Kalbfleisch (1986), Keifer (1988), Greene (1998), Güell and Hu (2006) Dixon and Le Bihan (2012), Chen, Diebold and Schorfheide (2013) and many others.

Evolution of the literature in duration analysis:

- Cox (1972), Lancaster (1979), Elbers and Ridder (1982), Orme (1989), Nickell (1979), Keifer and Neumann (1981), Kennan (1985), Staigler, Stock (1997)
- Lancaster (1990),Hey and Orme (1994) Greene (1998), Lancaster and Chesher (1983), Dobson and Goddard (2010)

- Chesher (1984), Greene (1998), Imbens and Lancaster (1994), Lancaster (1990), Struthers and Kalbfleisch (1986)
- Fasiani (1934), Lockwood (1991), Blanchard and Diamond (1994), McCall (1994), Zhang, Russell, Tsay (2001) Lalive, Van Ours and Zweimuller (2006), Dixon and Le Bihan (2012), Chen, Diebold and Schorfheide(2013), Hausman and Woutersen (2014)
- Grammig and Maurer (2000), Bover, Arellano and Bentolila (2002), Feng , Jiang, and Song (2004) Albrecht and Vroman (2005),Lalive, Van Ours and Zweimüller (2008), Brinch (2011), Candelon, Colletaz, Hurlin, and Tokpavi (2011) Christensen and Rudebusch (2012), Brinch (2011)

What is Duration Analysis?

- There are several economic questions in which the investigator is interested to know how long certain thing will last given that it has survived/existed for so long time.
- Duration of these events is a random variable that depends on chances and duration analysis aims to analyse what factors determine the length of duration of occurrence for period up to T period ($t \leqslant T$) or survival after period T ($t \geqslant T$)or what is probability of transition or the hazard rate between T and $T + \Delta$ period.
- Modelling duration has been used to determine the duration or probability of termination strikes, unemployment, marriage, disaster spells, heart attacks or many other ill-spells, likelihood of bankruptcy of a firm, technological breakthrough, probability of maintaining championship titles in sports).
- Main questions of these is to study if the event existed so far how long will it last or what is the rate of survival next period? For instance manager of a company would be interested to know how long will a certain machine last given that it has been running so far?
- A life insurance company would be interested in probability of death of an individual with certain medical record or physical characteristic in the next $T+\Delta$ years given that the person has survived up to T years. A union leader or the management negotiator will be interested about the probability of withdrawal of a strike given that the strike has continued up to T periods.

Example of Duration Analysis

- Main questions of these is to study if the event existed so far how long will it last or what is the rate of survival next period?
- For instance manager of a company would be interested to know how long will a certain machine last given that it has been running so far?
- A life insurance company would be interested in probability of death of an individual with certain medical record or physical characteristic in the next $T+\Delta$ years given that the person has survived up to T years.
- A union leader or the management negotiator will be interested about the probability of withdrawal of a strike given that the strike has continued up to T periods.

Duration Density The starting point of duration analysis is cumulative density function for duration which gives the distribution of duration variable starting from an initial state 0 up to period t as following:

$$\Pr\left(t \leqslant T\right) = F\left(t\right) = \int_0^t f\left(t\right)\partial t \tag{1475}$$

More interesting is the survival rate which is

$$S\left(t\right) = 1 - F\left(t\right) = \Pr\left(t > T\right) \tag{1476}$$

Probability of transition from one state to another (from unemployment of to employment, life to death, working condition to break down) is given by a hazard rate or probability of termination

Hazard Rate

Hazard rate

$$\lambda\left(t\right) = \lim_{\Delta\longrightarrow 0} \frac{F\left(t+\Delta\right) - F\left(t\right)}{\Delta S\left(t\right)} = \frac{f\left(t\right)}{s\left(t\right)} \tag{1477}$$

$$f\left(t\right) = s\left(t\right).\lambda\left(t\right) \tag{1478}$$

Hazard function is linked to the survival function as

$$\lambda\left(t\right) = \frac{\partial\log\left[1-F\left(t\right)\right]}{\partial t} = \frac{-F\left(t\right)}{1-F\left(t\right)} = \frac{f\left(t\right)}{s\left(t\right)} \tag{1479}$$

It is possible to derive the duration function by integrating the survival function

Duration and Survival Fuctions It is possible to derive the duration function by integrating the survival function

$$\int_0^t f\left(t\right)\partial t = -\partial\log\left[1-F\left(s\right)\right] + \log\left[1-F\left(0\right)\right] = -\log\left[1-F\left(s\right)\right] \tag{1480}$$

$$F\left(s\right) = 1 - \exp\left(\int_0^t \lambda\left(t\right)\partial t\right) \tag{1481}$$

Therefore modelling hazard function is the main element in duration models.

$$\lambda\left(t, x_i\right) = \lambda_0\left(t\right)\exp\left(x_i^{'}, \beta\right) \tag{1482}$$

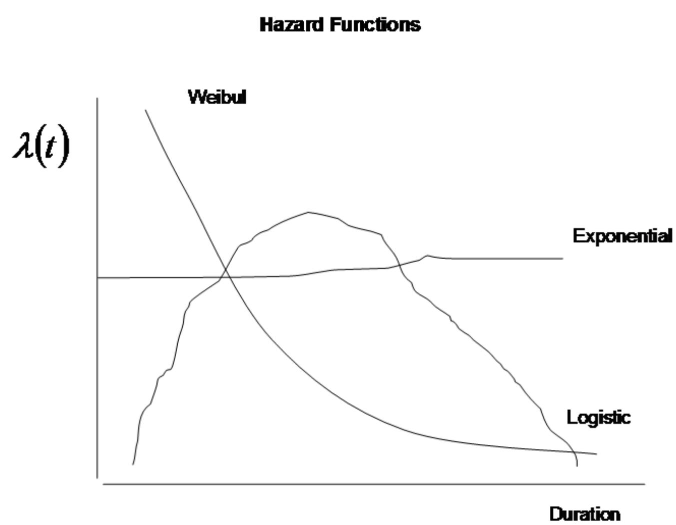
Main Points in Duration Analysis

- Important element is this is modelling the duration dependence, that give the likelihood of how much hazard rate depends on the duration variable.
- There is positive duration dependence if the longer the time spent in a given state, the higher the probability of leaving it soon.
- For instance, longer a light bulb works the higher the probability that it fails next period. Negative duration dependence implies longer the time spent in a given state, the lower the probability of leaving it soon.

- For instance, the longer the job search lasts, the less chance an unemployed person has finding a job.
- Absence of duration dependence is observed if the duration does not impact on the hazard rate, but this case is less appealing than the positive or negative duration dependence.
- Duration dependence $\frac{\lambda(t)}{\partial t} > 0$ indicates positive duration dependence and $\frac{\lambda(t)}{\partial t} < 0$ indicates negative duration dependence. Whereas $\frac{\lambda(t)}{\partial t} = 0$ indicates no duration dependence.
- There are a number of ways of modelling the hazard functions;

  1. exponential
  2. Weibull
  3. gamma
  4. log-normal and logistic hazard
  5. GAMMA

models are more popular in the literature (See Wooldridge (2002: chapter 20); Green (2008), Chap 25).



9.0.2 Exponential hazard model

Here T has exponential distribution. $1 - F(t) = 1 - \exp(-\lambda.t)$. This distribution does not have memory $\lambda(t) = \lambda$, the hazard rate does not depend on duration, it is constant $\lambda(t) = \lambda$.

$f(t) = \lambda \exp(-\lambda.t)$ for $(t > 0)$

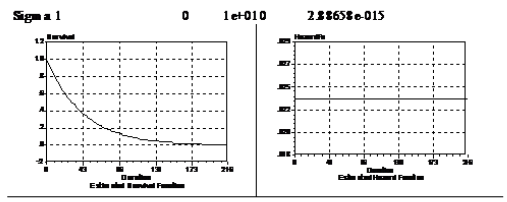
$$\lambda(t) = \frac{\partial \log [1 - F(t)]}{\partial t} = \frac{-\partial \log S(t)}{\partial t} \tag{1483}$$

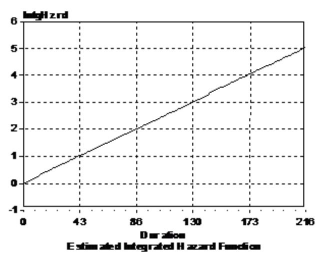
$$\ln S(t) = k - \lambda(t) = k - \lambda.t \tag{1484}$$

$$S(t) = K \exp(-\lambda t) \tag{1485}$$

Estimation of λ is simple; expected duration $E(t) = \frac{1}{\lambda}$ is and the maximum likelihood estimation of λ is .$\frac{1}{t}$

Integrated hazard function is written as $\Lambda(t) = \int_0^t \lambda(t)\, \partial t$ or $S(t) = \exp(-\Lambda(t))$ or $\Lambda(t) = -\ln S(t)$





Exponential hazard: an example
Cummulative density function (CDF) duration:

$$F(t) = 1 - e^{-\lambda t} \tag{1486}$$

Duration density:

$$f(t) = F'(t) = \lambda e^{-\lambda t} \tag{1487}$$

Survival function

$$S(t) = 1 - F(t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-\lambda t} \tag{1488}$$

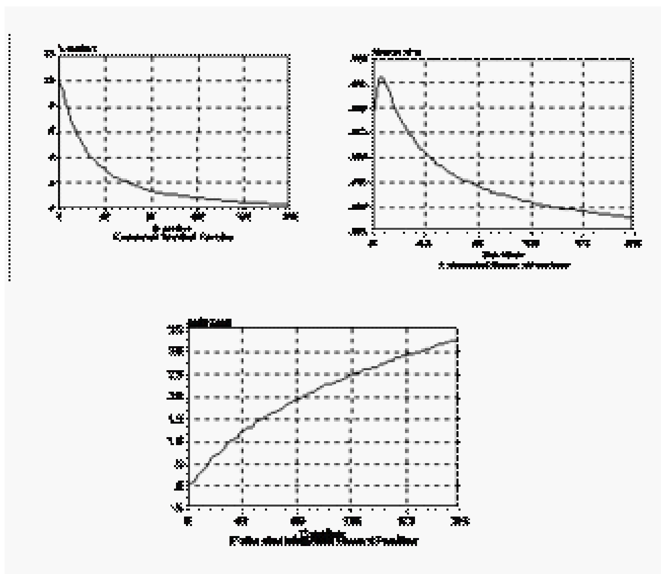
Hazard function:

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \tag{1489}$$

Thus the hazard rate λ is a constant in the exponential hazard model.
See examples in STATA and NLOGIT/LIMDEP: Spell, duration, aps, BHPS, recid_jw.

9.0.3 Weibull Hazard Model

- The CDF of T is given by $F(t) = 1 - \exp(-\lambda.t^{\alpha})$where λ and α are nonnegative parameters; and the density is given by
$$f(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha}) \tag{1490}$$
- and the hazard function is $.\lambda(t) = \frac{f(t)}{s(t)} = \frac{\alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha})}{\exp(-\lambda.t^{\alpha})} = \alpha\lambda t^{\alpha-1}$
- When $\alpha = 1$, the Weibull distribution reduces to the exponential distribution with $\lambda(t) = \lambda$; if $\alpha > 1$, the hazard is monotonically increasing, $\lambda(t) = \alpha\lambda t^{\alpha-1}$, which shows positive duration dependence. If $\alpha < 1$, the hazard , is continuously decreasing.$\lambda(t) = \alpha\lambda t^{\alpha-1}$
- Thus the Weibull distribution is better to capture the duration variable and transition between states if the hazard is monotonically increasing or decreasing.



9.0.4 Log Normal

- Log normal distributions of durations give non-monotonic hazard functions; first the hazard rate increases with duration and then decreases.
- This type of analysis is good in modelling bankruptcy rates. When follows a normal distribution with mean m and variance it follows the normal distribution; its density is given by following function

$$f(t) = \frac{1}{\sigma.t}\phi\left(\frac{\log T - m}{\sigma}\right) \tag{1491}$$

and the survivor function is $S(t) = 1 - \Phi\left(\frac{\log T - m}{\sigma}\right)$ with Φ denoting the CDF of a standard normal. The hazard function using

$\lambda(t) = \frac{f(t)}{s(t)}$

$$\lambda(t) = \frac{f(t)}{s(t)} = \frac{1}{T}\frac{\frac{1}{\sigma}\phi\left(\frac{\log T - m}{\sigma}\right)}{1 - \Phi\left(\frac{\log T - m}{\sigma}\right)} \tag{1492}$$

9.0.5 Log logistic Hazard Model

Log logistic hazard function is

$$F(t) = 1 - \frac{1}{(1 + \gamma t^{\alpha})}; \qquad S(t) = \frac{1}{(1 + \gamma t^{\alpha})}$$

$$\lambda(t) = \frac{f(t)}{s(t)} = \frac{\alpha\gamma t^{\alpha-1}}{(1 + \lambda t^{\alpha})^2} \div \frac{1}{(1 + \gamma t^{\alpha})} \times = \frac{\gamma\alpha t^{\alpha-1}}{1 + \gamma t^{\alpha}} \tag{1493}$$

where the α and γ are positive parameters.

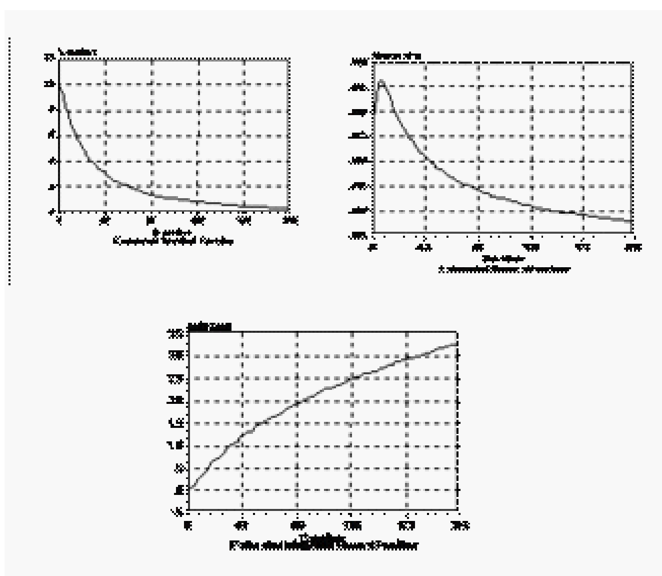
$$\int_0^{\infty} \lambda(st)\,\partial s = \int_0^{\infty} \frac{\gamma\alpha t^{\alpha-1}}{1 + \gamma t^{\alpha}}\partial s = \log(1 + \gamma t^{\alpha}) = -\left[\log(1 + \gamma t^{\alpha})^{-1}\right] \tag{1494}$$

Using $F(s) = 1 - \exp\left(\int_0^t \lambda(t)\,\partial t\right)$ condition derived above

$$F(t) = 1 - \frac{1}{(1 + \gamma t^{\alpha})} = 1 - (1 + \gamma t^{\alpha})^{-1} \quad for\ t > T \tag{1495}$$

Differentiating with respect to t gives:

$$f(t) = \alpha\gamma t^{\alpha-1}(1 + \lambda t^{\alpha})^{-2} \quad ; \ S(t) = \frac{1}{(1 + \gamma t^{\alpha})} \tag{1496}$$



GAMMA Hazard Model and Summary

$$f(t)=\frac{\left[a^{v}t^{v-1}\exp(-at)\right]}{\Gamma(v)}\quad \text{where}\ \ \Gamma(v)=\int_{0}^{\infty}\exp(-t)\,t^{v-1}\partial s \tag{1497}$$

Summary of popular distributions for duration model
Exponential functions for survival.
$S(t)=\exp(-\Lambda(t))\quad \lambda(t)=\lambda\quad F(t)=1-\exp(-\lambda.t)\quad f(t)=\lambda\exp(-\lambda.t)$
Logistic
$S(t)=1-\Phi\left(\frac{\log T-m}{\sigma}\right)\ \lambda(t)=\frac{1}{T}\frac{\frac{1}{\sigma}\phi\left(\frac{\log T-m}{\sigma}\right)}{1-\Phi\left(\frac{\log T-m}{\sigma}\right)}\ \ F(t)=1-(1+\alpha t^{\alpha})^{-1}\ \ f(t)=\alpha\gamma t^{\alpha-1}(1+\lambda t^{\alpha})^{-2}$
Weibull
;$S(t)=\exp(-\lambda.t^{\alpha})$; ; $\lambda(t)=\alpha\lambda t^{\alpha-1}\ \ F(t)=1-\exp(-\lambda.t^{\alpha})\quad f(t)=\alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha})$

9.1 Estimation of Hazard Models

Log linear models: Parameters of above models $\theta=(\lambda,\gamma)$ can be estimated using the maximum likelihood function for uncensored and censored observations.

$$\ln L=\sum\ln f(t/\theta)+\sum\ln s(t/\theta) \tag{1498}$$

It is easily estimated by BHHH (Berdt-Hall-Hall-Hauseman (1974) estimator (See Greene (938-951)).

$$\ln L=\sum\ln\lambda(t/\theta)+\sum\ln s(t/\theta) \tag{1499}$$

Proportional hazard models,

$$\lambda(t)=e^{-\beta(t,\theta)}\lambda_{0}(t_{i}) \tag{1500}$$

where the $\lambda(t)$ is proportional to the baseline hazard function $\lambda_0(t_i)$.
Empirical implementation (Greene (2000); Chapter 20; Using Limdep)

9.1.1 Estimation of Duration in STATA

See the log file hazard and hazard1 from the Annual Population Survey

*apsp_jul07jun08_rw09_260310.dta

use "Z:\Econometrics\programs\STATA\apsp_jul07jun08_rw09_260310.dta", clear

stset durun

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(weibull)

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(exponential)

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(gompertz)

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(lognormal)

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(loglogistic) streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99 grsexp, dist(weibull)

failure _d: 1 (meaning all fail)

analysis time _t: durun

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99

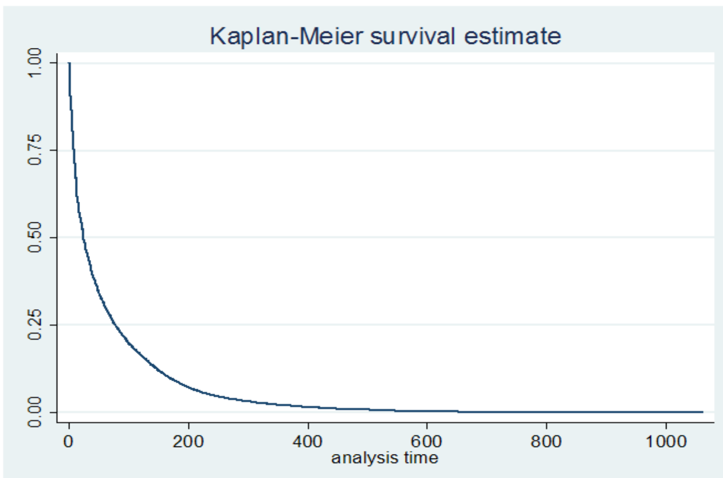
> grsexp, dist(exponential)

streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99

> grsexp, dist(gompertz)

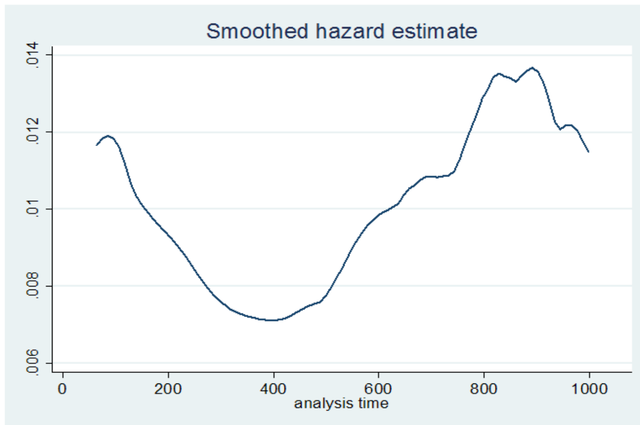
streg tpben31 tpben32 tpben33 tpben34 tpben35 tpben36 self1 self2 self3 self4 sex ethas ethbl gross99

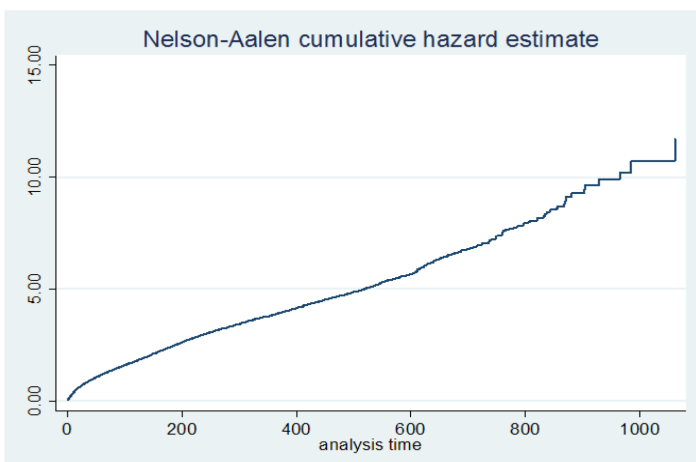
> grsexp, dist(lognormal)



Estimation of Duration in STATA

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
tpben31	.9708788	.0020447	-14.03	0.000	.9668796	.9748946
tpben32	.9733548	.0020842	-12.61	0.000	.9692784	.9774483
tpben33	1.004645	.0029922	1.56	0.120	.9987971	1.010527
tpben34	.9886277	.0086266	-1.31	0.190	.9718636	1.005681
tpben35	.9717948	.0358969	-0.77	0.439	.9039247	1.044761
tpben36	(omitted)					
self1	1.020205	.0044089	4.63	0.000	1.0116	1.028883
self2	.9845701	.0131169	-1.17	0.243	.9591941	1.010617
self3	.9888829	.0183726	-0.60	0.547	.9535209	1.025556
self4	1.028848	.0244766	1.20	0.232	.9819762	1.077957


Smoothed hazard estimate
.006
.008
.01
.012
.014
0
200
400
600
800
1000
analysis time


Nelson-Aalen cumulative hazard estimate
0.00
5.00
10.00
15.00
0
200
400
600
800
1000
analysis time

sex | 1.531057 .0346324 18.83 0.000 1.464662 1.600462
ethas | .9877391 .0038554 -3.16 0.002 .9802115 .9953245
ethbl | .9862598 .0051397 -2.65 0.008 .9762374 .9963851
gross99 | (omitted)
grsexp | (omitted)
————-+————————————————————————-
/ln_p | .474606 .0080713 58.80 0.000 .4587865 .4904255
————-+————————————————————————-
p | 1.607381 .0129737 1.582153 1.633011
1/p | .6221301 .0050214 .6123658 .6320502

NLogit (LIMDEP) Commands: NLOGIOT by William Green is special software for Cross Section and Duration Analysis

```
/*==========================================================================
Example 20.17. Log-Linear Survival Models for Strike Duration
*/==========================================================================
Read ; Nobs = 62 ; Nvar = 2 ; Names = T,Prod $
T Prod
7    0.01138
14   0.01138
52   0.01138
37   0.02299

? Four survival models for duration
?
Create ; logT = Log(T) $
Surv; Lhs=logT ; Rhs = One ; Model=Exponential ; Plot$
Surv; Lhs=logT ; Rhs = One ; Model=Weibull ; Plot$
Surv; Lhs=logT ; Rhs = One ; Model=Logistic ; Plot$
Surv; Lhs=logT ; Rhs = One ; Model=Normal ; Plot $
```

Original Research Article

References

[1] Albrecht, J. and Vroman, S. (2005), Equilibrium Search With Time-Varying Unemployment Benefits. The Economic Journal, 115: 631–648

[2] Blanchard O. J. and Peter Diamond (1994) Ranking, Unemployment Duration, and Wages, Review of Economic Studies 61 (3): 417-434

[3] Bover, O., Arellano, M. and Bentolila, S. (2002), Unemployment Duration, Benefit Duration And The Business Cycle. Economic Journal, 112: 223–265.

[4] Brinch, C. N. (2011), Non-parametric identification of the mixed proportional hazards model with interval-censored durations, Econometrics Journal, 14: 343–350.

[5] Burgess, S. and Turon, H. (2005), Unemployment dynamics in Britain, Economic Journal, 115: 423–448.

[6] Candelon B, Gilbert Colletaz, Christophe Hurlin, and Sessi Tokpavi (2011) Backtesting Value-at-Risk: A GMM Duration-Based Test, Journal of Financial Econometrics 9 (2): 314-343

[7] Chen F, F. X. Diebold and F Schorfheide(2013) A Markov-switching multifractal inter-trade duration model with application to US equities, Journal of Econometrics, 177, 2, 320-342

[8] Chesher A (1984) Improving the efficiency of Probit estimators, Review of Economic Studies,66:3:523-527.

[9] Christensen, J. H. E. and Rudebusch, G. D. (2012), The Response of Interest Rates to US and UK Quantitative Easing, Economic Journal, 122: F385–F414.

[10] Cox D. R. (1972) Regression Models and Life-Tables, Journal of the Royal Statistical Society, B 34,2, 187-220

[11] Dixon H and H. Le Bihan (2012) Generalised Taylor And Generalised Calvo Price And Wage Setting: Micro-Evidence With Macro Implications, Economic Journal, 122 (May), 532–554

[12] Dobson S and JA Goddard (2010) Optimizing strategic behaviour in a dynamic setting in professional team sports, European Journal of Operational Research, 205:3 661-669.

[13] Elbers C. and G. Ridder (1982) True and spurious duration dependence: the identifiability of proportional hazard model, Review of Economic Studies, 49:3:July: 403-409

[14] Fasiani M (1934) Materials for a Theory of the Duration of the Process of Tax Shifting, Review of Economic Studies, 1 (2): 81-101

[15] Feng D, G. J. Jiang, and P. X.-K. Song (2004) Stochastic Conditional Duration Models with "Leverage Effect" for Financial Transaction Data, Journal Of Financial Econometrics, 2 (3): 390-421

[16] Grammig, J. and Maurer, K.-O. (2000), Non-monotonic hazard functions and the autoregressive conditional duration model, Econometrics Journal, 3: 16–38.

[17] Greene W. (2008) Econometric Analysis, Prentice Hall, 6th edition.

[18] Greene W.H. (2007) NLOGIT4 (LIMDEP Version 10) User Manual, Econometric Software Inc, New York.

[19] Gujarati D. (2011) Econometrics by Examples, Palgrav macmillan.

[20] Güell M and L Hu (2006) Estimating the probability of leaving unemployment using uncompleted spells from repeated cross-section data, Journal of Econometrics,133,1, 307-341

[21] Hausman J. A. and T. Woutersen (2014) Estimating a semi-parametric duration model without specifying heterogeneity, Journal of Econometrics, 178, Part 1, 114-131

[22] Hausman J.A., (1978), Specification Tests in Econometrics, Econometrica, 46 : 6, pp.1251-1271.

[23] Heckman J. J., (1979), Sample Selection Bias as a Specification Error, Econometrica, 47 : 1, pp153-161.

[24] Heckman J. J. and B. Singer (1984) Econometric duration analysis, Journal of Econometrics, 24, 1–2, 63-132

[25] Hey J. D.and C. Orme (1994) Investigating Generalizations of Expected Utility Theory Using Experimental Data, Econometrica, 62, 6, 1291-1326.

[26] Huw D. and H. Le Bihan (2012) Generalized Taylor and Generalized Calvo price and wage-setting: micro evidence with macro implications, Economic Journal, 122 (May), 532–554.

[27] Imbens G. W. and T Lancaster (1994) Combining Micro and Macro Data in Microeconometric Models, Review of Economic Studies, 61:4:655-680.

[28] Kennan J (1985) The duration of contract strikes in US Manufacturing, Journal of Econometrics 28 (1985) 5-28.

[29] Keifer N (1988) Economic duration data and hazard functions, Journal of Economic Literature, 26:647-679.

[30] Keifer,N . M. and Neumann,G . R. (1981), "Individual effects in a Non-Linear Model:Explicit Treatment of Heterogeneity in the Empirical Job-Search Model", Econometrica 4,9, 965-979.

[31] Lancaster T (1979) Econometric Methods for Duration of Unemployment, *Econometrica*, 47:4:939-56.

[32] Lancaster T (1990) Econometric Analysis of Transition Data, Blackwell

[33] Lancaster T and A Chesher (1983) The Estimation of Models of Labour Market Behviour, Review of Economic Studies, 50:4:609-624.

[34] Lalive R (2008) How do extended benefits affect unemployment duration? A regression discontinuity approach Journal of Econometrics, 142, 2, 785-806

[35] Lalive R, Jan Van Ours and Josef Zweimuller (2006) How Changes in Financial Incentives Affect the Duration of Unemployment, Review of Economic Studies 73 (4): 1009-1038

[36] Lalive, R., Van Ours, J. C. and Zweimüller, J. (2008), The Impact of Active Labour Market Programmes on The Duration of Unemployment in Switzerland, Economic Journal, 118: 235–257.

[37] Lockwood B (1991) Information Externalities in the Labour Market and the Duration of Unemployment, Review of Economic Studies , 58 (4): 733-753

[38] McCall B.P (1994) Specification diagnostics for duration models: A martingale approach, Journal of Econometrics, 60, 1–2, 293-312

[39] Nickell S.J.(1979) Estimatingt he Probability of Leaving Unemployment, Econometrica, 4, 7, 1249-1264

[40] Ohn, J., Taylor, L. W. and Pagan, A. (2004), Testing for duration dependence in economic cycles, Econometrics Journal, 7: 528–549

[41] Orme C. (1989) On the uniqueness of the maximum likelihood estimator in the truncated regression models, Econometric Review, 8:2:217-222.

[42] Staigler D., Stock J. H., (1997), Instrumental Variables Regression with Weak Instruments, Econometrica, 65: 3, pp.557-586.

[43] Struthers C. A. and J. D. Kalbfleisch (1986) Misspecified Proportional Hazard Models, Biometrica, 73, 2, 363-369

[44] Verbeek M. (2012) A Guide to Modern Econometrics, Wiley 3rd edition.

[45] Wooldridge J. M. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press.

[46] Zhang M.Y., J. R. Russell, R. S. Tsay (2001) A nonlinear autoregressive conditional duration model with applications to financial transaction data, Journal of Econometrics, 104, 1, 179-207

9.1.2 Tutorial 5: Duration Analysis

Q1. Derive duration density, hazard rate, survival function and duration dependence for the following duration or hazard functions and explain the general procedure for estimation of model parameters.

a. Exponential distribution. $[F(t) = 1 - \exp(-\lambda.t)]$.

b. Weibull $\left[f(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha})\right]$.

c. Log normal distribution $\left[f(t) = \frac{1}{\sigma.t}\phi\left(\frac{\log T - m}{\sigma}\right)\right]$.

d. Log logistic $\left[f(t) = \alpha\gamma t^{\alpha-1}(1+\gamma t^{\alpha})^{-2}\right]$.

e. Gamma distribution $\left[f(t) = \frac{\left[a^{v}t^{v-1}\exp(-at)\right]}{\Gamma(v)} \quad \text{where} \quad \Gamma(v) = \int_0^{\infty}\exp(-t)\, t^{v-1}\partial s.\right]$

10 L9: Bayesian Analysis

True parameters of a model are unknown both in the classical and Bayesian models but there is one major difference. True parameters have a given density function in the classical models but such density keeps changing when new information arrives in the Bayesian model. Due to parameter uncertainty the Bayesian model is becoming increasingly popular in recent years. **Bayesian methods were developed in** Lancaster(1979), Lancaster and Chesher (1983), Imbens and Lancaster (1994), Bauwens, Lubrano and Richard (1999), Koop (2003), Anscombe (1961),Pratt (1965), Doan, Litterman and Sims (1984), Berger (1990), Chib (1993), Rust (1996) Phillips and Ploberger (1996) Bauwens, Lubrano and Richard (1999) Judge,Griffiths, Hill, Lutkepohl and Lee (1990) Geweke and Keane (2000), Chib, Nardarib and Shephard (2002), Heckelei and Mittelhammer (2003) Canova and Ciccarelli (2004),George, Sun and Ni (2008), Levine, Pearlman, Perendia and Yang (2012).

Thusd the literature in Bayesian Econometrics:

- Bayes (1763), Lancaster(1979), Lancaster and Chesher (1983), Imbens and Lancaster (1994), Bauwens, Lubrano and Richard (1999), Koop (2003), Pratt (1965), Doan, Litterman and Sims (1984), Berger (1990), Chib (1993), Rust (1996) Phillips and Ploberger (1996) Bauwens, Lubrano and Richard (1999) Judge,Griffiths, Hill, Lutkepohl and Lee (1990) Geweke and Keane (2000), Chib, Nardarib and Shephard (2002), Heckelei and Mittelhammer (2003)

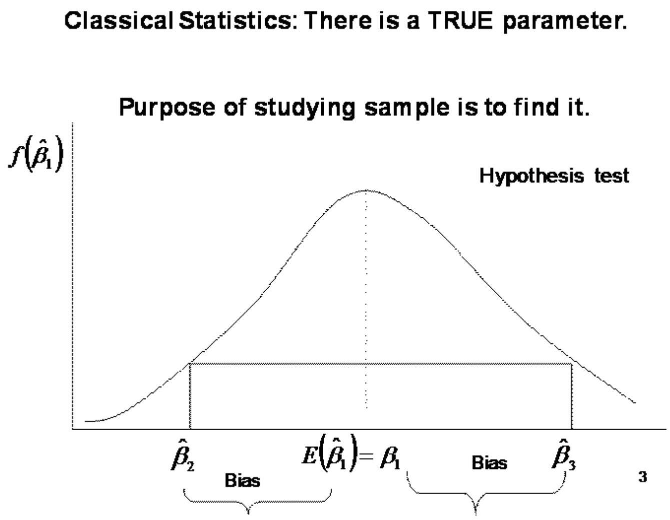
- Anscombe (1961), Chamberlain (2000), Chib (1993), Chib and Greenberg (1994)
- Chib, Nardarib and Shephardc (2002), Lancaster (2004), Chib , Greenberg, Winkelmann (1998)
- Geweke and Keane (2000) Harrison and Stevens (1976), Heckman and Macurdy (1980)
- Kleibergen and Zivot (2003) Koop and Van Dijk (2000) Lancaster (1997)
- Rossi and Allenby (2003) Schotman and van Dijk (1991), Sims and Uhlig (1991),
- Zellner (1985), Koop (1992), Phillips and Ploberger (1996), Pesaran and Smith (1996), Sims and Zha (1998), Rust (1996), Litterman (1986) Heckelei and Mittelhammer (2003), Canova and Ciccarelli (2004)
- Levine, Pearlman, Perendia and Yang (2012)
- Chetty and Sankar (1969), Geweke(1996), Geweke (2001), Li , Zeng and Yu (2014), Conley, Hansen, McCulloch and Rossi (2008), Musalem, Bradlow and Raju (2009), Bańbura, Giannone and Reichlin (2010) , Carriero, Kapetanios and Marcellino (2011), Norets and Tang (2013), Sala (2014)

Bayesian analysis is parallel to the classical analysis as models with single or multiple equations, simultaneous equations, VAR, Panel or time varying parameters each could be estimated in a Bayesian way.

- Bayesian econometrics is an approach to estimate parameters and models involved in economic analysis which is different from the classical approach.
- Classical econometric methods assume that there is true parameter underlying the data generating process such as θ and its true value is unknown. The objective of sample statistics $\widehat{\theta}$ is to represent this unknown parameter as best as possible.
- The estimated parameter $\widehat{\theta}$ is a random variable and has its own distribution where as the true parameter θ is a fixed number but unknown.
- The estimator should be unbiased, $\left[E\left(\widehat{\theta}\right)-\theta\right]=0$, and the efficiency of an estimator is judged by the minimum square error $\left[E\left(\widehat{\theta}\right)-\theta\right]\left[E\left(\widehat{\theta}\right)-\theta\right]^{\prime}$ and the data generating process is given by

$$f\left(y;\theta\right)=\Pi\frac{1}{\sqrt{2\pi\sigma^{2}}}\exp\left[\frac{-\left(y-\mu\right)^{2}}{2\sigma^{2}}\right] \tag{1501}$$

Classical Econometrics



10.1 What is Bayesian Analysis?

- Analysis in classical econometrics is based on an assumption that true population parameters are constant but sample estimates of those parameters are random variables distributed normally around those population parameters.
- Bayesian regard true parameter to be a random variable, prior are updated freqeuntly upon arrival of new data.

In Bayesian analysis the value of true parameter θ unknown like in the classical approach but it is not fixed.

Instead θ has a probability distribution and it is updated continuously based on sample information – priors.

The prior density is given by $f(\theta)$ and this may represent all available information up to that point.

$$f(\theta_1) = \int_{-\infty}^{\infty} f(\theta)\, d\theta_2 d\theta_3 ... d\theta_n \tag{1502}$$

The sample density of variable y is treated as conditional on the random variable θ given by $f(y/\theta)$.

The joint density of y and θ is result of the product of prior density and the sample likelihood function.

Components of the Bayesian Model A Bayesian econometric model

$$f(\theta/y) = \frac{f(y/\theta)\, f(\theta)}{f(y)} \tag{1503}$$

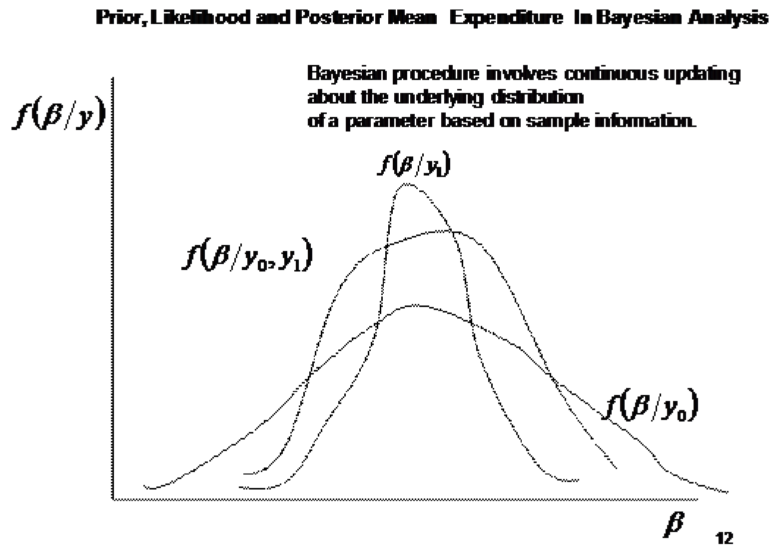
has two components;

$f\left(y/\theta\right)$ is called the likelihood function, it describes what you see for every particular values of parameter set $\theta\in\Theta$ as the in the consumption function with mean and variance $\theta_1=(10,0.9)$

The second component $f\left(\theta\right)$ is the distribution over the parameter space Θ; prior distribution denoting the beliefs about particular parameter values $\theta\in\Theta$.

Parameter $\theta\in\Theta$ is unknown both before and after the data has been observed but data are unknown before the information has been gathered but known afterwards.

$f\left(\theta/y\right)$ is the posterior formed after taking account of prior and data.



Bayesian algorithm (Lancaster (2004))

1. Formulate an economic model with conditional probability distribution over parameter space $\theta\in\Theta$; such as $f\left(\theta/y\right)=\frac{f(y/\theta)f(\theta)}{f(y)}$
2. Organise beliefs about θ into a prior probability distribution over Θ.
3. Collect data and insert into the model as given in step 1.
4. Criticise the model.

10.2 Bayesian Rule

Let $p(A,B)$ be the joint probability of occuring events A and B together, $p\left(B\right)$ be the marginal probability of B without any respect to occurence of A then the probability of A conditional on the occurance of B is

$$p\left(A|B\right)=\frac{p(A,B)}{p\left(B\right)} \tag{1504}$$

Similarly the probability of B conditional on the occurance of A is

$$p\,(B|A) = \frac{p(A,B)}{p\,(A)} \tag{1505}$$

Substituting the value of $p(A,B)$ from 1 the probability of B conditional on the occurance of A is

$$p\,(B|A) = \frac{p\,(A|B)\,p\,(B)}{p\,(A)} \tag{1506}$$

Bayesian econometrics is application of Bayesian Rule repeatedly for estimation on unknown parameters.

For data y and parameter θ then by replacing A by y and B by θ

$$p\,(\theta|y) = \frac{p\,(y|\theta)\,p\,(\theta)}{p\,(y)} \tag{1507}$$

Since $p\,(y)$ does not involve θ it can be ignored and this function written as

$$p\,(\theta|y)\ \propto p\,(y|\theta)\,p\,(\theta) \tag{1508}$$

where $p\,(\theta|y)$ is the posterior density, $p\,(y|\theta)$ is the likelihood function and $p\,(\theta)$ is the prior density. Posterior thus is proportional to likelihood times prior. Posterior combines both data and non-data information.

10.3 Bayesian Likelihood

Consider a likelihood function in the classical analysis

$$L(\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_i^N (Y-\beta X)^2\right] \tag{1509}$$

$$L(\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_i^N \left(Y-\widehat{\beta}X-\beta X+\widehat{\beta}X\right)^2\right] \tag{1510}$$

$$L(\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_i^N \left(Y-\widehat{\beta}X-\beta X+\widehat{\beta}X\right)^2\right] \tag{1511}$$

$$L(\beta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp-\frac{1}{2\sigma^2}\left[\sum_i^N (Y-\widehat{\beta}X)^2-\left(\beta-\widehat{\beta}\right)^2\sum_i^N X_i^2\right] \tag{1512}$$

$\frac{\sum(Y-\widehat{\beta}X)^2}{N-1} = s^2$ with $v = N-1$

$$\sum_i^N (Y-\beta X)^2 = vs^2+\left(\beta-\widehat{\beta}\right)^2\sum_i^N X_i^2 \tag{1513}$$

$$L(\beta,\sigma^2) = \frac{1}{(2\pi)^{\frac{N}{2}}}\left\{\frac{1}{\sqrt{\sigma^2}}\exp\left[-\frac{1}{2}\frac{\left(\beta-\widehat{\beta}\right)^2}{\sigma^2\left(\sum_i^N X_i^2\right)^{-1}}\right]\frac{1}{\left(\sqrt{\sigma^2}\right)^v}\exp\left[-\frac{vs^2}{2\sigma^2}\right]\right\} \tag{1514}$$

Priors and Posteriors

$$L(\beta,\sigma^2) \propto \frac{1}{\sqrt{\sigma^2}}\exp\left[-\frac{1}{2}\frac{\left(\beta-\widehat{\beta}\right)^2}{\sigma^2\left(\sum_i^N X_i^2\right)^{-1}}\right] \tag{1515}$$

Prior hyperparameters (prior elicitation)

$$\beta \sim N\left(\underline{\beta}, \sigma^2\underline{V}\right) \tag{1516}$$

Posterior

$$p(\beta|y) \propto p(y|\beta)\,p(\beta) \tag{1517}$$

Priors and Posteriors

$$p(\beta|y) \propto p(y|\beta)\,p(\beta) \propto \left\{\frac{1}{\sqrt{\sigma^2}}\exp\left[-\frac{1}{2}\frac{\left(\beta-\widehat{\beta}\right)^2}{\sigma^2\left(\sum_i^N X_i^2\right)^{-1}}\right]\frac{1}{\sqrt{\sigma^2\underline{V}}}\exp\left[-\frac{1}{2}\frac{\left(\beta-\underline{\beta}\right)^2}{\sqrt{\sigma^2\underline{V}}}\right]\right\} \tag{1518}$$

$$\propto \frac{1}{\sqrt{\sigma^2}}\exp\left[-\frac{1}{2}\frac{\left(\beta-\widehat{\beta}\right)^2}{\sigma^2\left(\sum_i^N X_i^2\right)^{-1}} - \frac{1}{2}\frac{\left(\beta-\underline{\beta}\right)^2}{\sqrt{\sigma^2\underline{V}}}\right] \tag{1519}$$

$$\propto \exp\left[-\frac{1}{2}\frac{\left(\beta-\overline{\beta}\right)^2}{\sigma^2\overline{V}}\right] \tag{1520}$$

Bayesian posterior

$$\overline{V} = \frac{1}{\underline{V}^{-1} + \sum_{i}^{N} X_i^2} \tag{1521}$$

$$\overline{\beta} = \overline{V}\left(\underline{V}^{-1} + \underline{\beta}\sum_{i}^{N} X_i^2\right) \tag{1522}$$

Bayesian posterior

$$\beta/y \sim N\left(\overline{\beta}, \sigma^2 \overline{V}\right) \tag{1523}$$

This has obvious parallel to the classical regression estimate. Extreme bound analysis, non-informative priors, relatively non-informative priors are other concepts that deserve some attention.

10.4 Bayesian Regression: Vague Prior

$$Y = X\beta + \varepsilon \tag{1524}$$

Vague prior

$$p(\beta, \tau) \propto \frac{1}{\tau}; \quad -\infty < \beta < \infty; \; \tau > 0 \tag{1525}$$

Posterior implied by the vague prior

$$p(\beta, \tau/y, X) \propto \tau^{\frac{n}{2-1}} \exp\left\{\left(-\frac{\tau}{2}\right)(Y - X\beta)'(Y - X\beta)\right\} \tag{1526}$$

joint posterior can be written as

$$p(\beta, \tau/y, X) \propto \tau^{\frac{n}{2-1}} \exp\left\{\left(-\frac{\tau}{2}\right)(\beta - b)X'X(\beta - b)\right\} \times \exp\left\{-\frac{\tau}{2}e'e\right\} \tag{1527}$$

Here $b = (X'X)^{-1} X'Y$ is the least square estimate and $e = Y - bX$ the least square residual. τ is precision parameter

Bayesian Regression: Marginal densities

marginal density of τ

$$p(\tau/y, X) \propto \tau^{\frac{n}{2-1}} \exp\left\{-\frac{\tau v s^2}{2}\right\}; \; s^2 = \frac{(Y - bX)'(Y - bX)}{N - K} = \frac{e'e}{v} \tag{1528}$$

marginal density of β in the linear model with vague prior

$$p(\beta/y, X) \propto \left[\left\{(\beta - b)' X'X(\beta - b)\right\} + vs^2\right]^{\frac{v+k}{2}} \tag{1529}$$

$$\propto \left[1 + \frac{(\beta - b)' X'X(\beta - b)}{vs^2}\right]^{\frac{v+k}{2}} \tag{1530}$$

$$\beta \sim t\left(b, s^2\left(X'X\right)^{-1}, v\right); \quad s^2 = \frac{(Y-bX)'(Y-bX)}{N-K} = \frac{e'e}{N-K} \tag{1531}$$

$$\beta_j \sim t\left(b_j, s^2\left(X'X\right)^{-1}_{i,j}, v\right); \quad t_v = \frac{\left(\beta_j - b_j\right)}{sd_j}; \quad sd_j = s\sqrt{\left(X'X\right)^{-1}_{i,j}} \tag{1532}$$

Bayesian Regression: Posterior moments of errors

Posterior elements of ε

Prior mean of ε is zero, posterior mean is e - the least square residual, posterior covariance matrix is $s^2X\left(X'X\right)^{-1}X'$

$$E(\varepsilon/y, X) = E(Y - X\beta/y, X) = y - XE(\beta/y, X) = y - Xb = e \tag{1533}$$

$$V(\varepsilon/y, X) = V(Y - X\beta/y, X) = y - XV(\beta/y, X) = s^2X\left(X'X\right)^{-1}X' \tag{1534}$$

10.4.1 Bayesian Regression: An Example

A joint posterior distribution could be approximated by the multivariate normal distribution with mean equal to ML estimate and precision eqaul to the negative hessain matrix at that estimate $\theta \sim N\left(\widehat{\theta}, -H\left(\widehat{\theta}\right)\right)$

$$\log l(\beta, \tau) = \frac{1}{2}\left[n\log\tau - \tau(\beta - b)X'X(\beta - b) - \tau e'e\right] \tag{1535}$$

Here $b = \left(X'X\right)^{-1}X'Y$ is the least seuare estimate and $e = Y - bX$ the least square residual. τ is precision parameter

First order conditions for maximising the log-likelihood

$$\frac{\partial \log l(\beta, \tau)}{\partial\beta} = -\tau(\beta - b)X'X = 0 \Longrightarrow \beta = b \tag{1536}$$

$$\frac{\partial \log l(\beta, \tau)}{\partial\tau} = \frac{n}{2\tau} - \frac{1}{2}\left\{(\beta - b)'X'X(\beta - b) + e'e\right\} = 0 \Longrightarrow \tau = \frac{n}{e'e} \tag{1537}$$

Bayesian Regression: Infomation Matrix (Hessian)

Taking the second difference

$$\frac{\partial^2 \log l(\beta, \tau)}{\partial^2\beta} = -\tau X'X; \quad \frac{\partial^2 \log l(\beta, \tau)}{\partial\beta\partial\tau} = -(\beta - b)X'X \tag{1538}$$

$$\frac{\partial^2 \log l(\beta, \tau)}{\partial^2\tau} = -\frac{n}{2\tau^2}; \quad \frac{\partial^2 \log l(\beta, \tau)}{\partial\tau\partial\beta} = -(\beta - b)'X'X \tag{1539}$$

$$-H\left(\widehat{\theta}\right) = \begin{bmatrix} \tau X'X & (\beta - b)X'X \\ (\beta - b)'X'X & -\frac{n}{2\tau^2} \end{bmatrix} \tag{1540}$$

Evaluated at the solution of the first order conditions

$$-H\left(\widehat{\theta}\right)=\left[\begin{array}{cc}\tau X'X & 0\\ 0 & -\frac{n}{2\tau^2}\end{array}\right] \tag{1541}$$

Bayesian Regression: Distribution of Posterior

$$\left[\begin{array}{c}\beta\\ \tau\end{array}\right]\sim N\left(\left[\begin{array}{c}b\\ \frac{n}{e'e}\end{array}\right],\left[\begin{array}{cc}\frac{n}{e'e}X'X & 0\\ 0 & \frac{\left(e'e\right)^2}{2n}\end{array}\right]\right) \tag{1542}$$

Here β and τ are approximately normally distributed with mean equal to the least square estimate b and the reciproval of the residual mean square. These parameters are independent and the precision depends only on n, $X'X$ and sum square of errors $(e'e)$

In large samples situations frequentist least square estimates and their standard deviations will be same as the Bayesian posterior means and standard deviations.

Bayesian approach can be applied to ARCH, GARCH, AR, MA or ARMA, Panel Data, VAR, Simultaneous equation model, cross section, count and duration models; unit root and cointegration and other time series models.

Next step is to check the posterior against beliefs or priors based on economic theory.

10.4.2 Bayesian Regression: Model Choice

Model choice based on Bayes' factor

$$B_{12}=\frac{p\left(y/1\right)}{p\left(y/2\right)} \tag{1543}$$

$$BIC=\left(\frac{l\left(\widehat{\theta}_1;y\right)}{l\left(\widehat{\theta}_2;y\right)}\right)n^{\frac{k_2-k_1}{2}}=\left(\frac{e_2'e_2}{e_1'e_1}\right)^{\frac{n}{2}}n^{\frac{k_2-k_1}{2}}=\left(\frac{1-R_2^2}{1-R_1^2}\right)^{\frac{n}{2}}n^{\frac{k_2-k_1}{2}} \tag{1544}$$

If two models give the idential residual square $(e_2'e_2)$ and if the model 2 has more parameters $(k_2>k_1)$ then choose the first model on the parsimonous ground

If two models have the same number of parameters then choose the model with small residual sum square.

Bayesian VAR: Theil's Mixed Estimation Method for Bayesian VAR Data generating process

$$\underset{T\times 1}{Y_j}=\underset{T\times K}{X_j}\underset{K\times 1}{\beta_j}+\underset{T\times 1}{u_j}\quad\text{where } u_j\sim N\left(0,\sigma_{i,j}^2 I_T\right) \tag{1545}$$

Prior distributions are

$$\underset{T\times 1}{r_j}=\underset{T\times K}{R_j}\underset{K\times 1}{\beta_j}+\underset{T\times 1}{v_j} \tag{1546}$$

Mixed estimator is obtained by combining above two

$$\left[\begin{array}{c}Y_j\\ r_j\end{array}\right]=\left[\begin{array}{c}X_j\\ R_j\end{array}\right]\left[\beta_j\right]+\left[\begin{array}{c}u_j\\ v_j\end{array}\right] \tag{1547}$$

Bayesian VAR: Theil's Mixed Estimation Method for Bayesian VAR

$$E\left[\begin{pmatrix} u_j & v_j \end{pmatrix}\begin{matrix} u_j \\ v_j \end{matrix}\right]=\begin{pmatrix} \sigma_{i,j}^2 I_T & 0 \\ 0 & I_T \end{pmatrix} \tag{1548}$$

Point estimation.

$$\beta_j=\left[\begin{pmatrix} X_j & R_j \end{pmatrix}\begin{pmatrix} \sigma_{i,j}^2 I_T & 0 \\ 0 & I_T \end{pmatrix}^{-1}\begin{pmatrix} X_j \\ R_j \end{pmatrix}\right]^{-1}\left[\begin{pmatrix} X_j & R_j \end{pmatrix}\begin{pmatrix} \sigma_{i,j}^2 I_T & 0 \\ 0 & I_T \end{pmatrix}^{-1}\begin{pmatrix} Y_j \\ r_j \end{pmatrix}\right]$$

10.4.3 Bayesian Panel Data Model

Bayesian panel equations look very similar to those in frequentist panel model

$$y_{i,t}=x_{i,t}\beta+\alpha_i+e_{i,t} \tag{1549}$$

with priors $e_{i,t}/x_{i,t},\beta,\alpha_i,\tau\sim N(0,\tau)$

Panel models allow to isolate individual specific effects α_i from other errors $e_{i,t}$. Here $e_{i,t}$ are independently distributed with mean zero and precision τ. In matrix notation

$$\underset{NT\times 1}{y_i}=\underset{NT\times(N+K)}{x_i}\underset{(N+K)\times 1}{\beta}+\alpha_i j_T+\underset{NT\times 1}{e_i}\quad where\ \ e_{i,j}\sim N\left(0,\sigma_{i,j}^2 I_T\right) \tag{1550}$$

$$y=\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_N \end{bmatrix}=\begin{bmatrix} x_1 & j_T & . & . & 0 & 0 \\ x_2 & 0 & j_T & . & 0 & 0 \\ . & . & 0 & j_T & 0 & 0 \\ . & . & . & . & j_T & 0 \\ x_N & 0 & . & . & 0 & j_T \end{bmatrix}\begin{matrix} \beta \\ \alpha_1 \\ \alpha_2 \\ . \\ \alpha_N \end{matrix}+\begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_N \end{bmatrix} \tag{1551}$$

$$Y = \left[\begin{pmatrix} X & R \end{pmatrix}\begin{pmatrix} \beta \\ \alpha \end{pmatrix}\right]+e\ \text{ for } R=[I_N\otimes j_T] \tag{1552}$$

$$Y = z\delta+e\ ;\ \text{e/z, } \delta\ \tilde{}\ \text{N}(0,\tau I_{NT}) \tag{1553}$$

Marginal distribution of error vector is normal, homoscedastic and non-autocorrelated.

$$\log l(\beta,\alpha,\tau;y,Z)\propto\tau^{\frac{NT}{2}}\exp\{-\tau(\delta-d)Z'Z(\delta-d)/2\}\times\exp\left\{-\frac{\tau}{2}e'e\right\} \tag{1554}$$

where $d=(Z'Z)^{-1}Z'Y$ is the least seuare estimate and $e=Y-Zd$

Univorm prior on $\delta=(\beta,\alpha)$

$$p(\beta,\alpha,\tau)\propto p(\delta,\tau)\ \propto\frac{1}{\tau};\quad -\infty<\beta<\infty;\ \tau>0 \tag{1555}$$

$\{\alpha_i\}$ is IID

$$\begin{bmatrix} X'X & X'R \\ R'X & R'R \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} X'Y \\ R'Y \end{bmatrix} \tag{1556}$$

$R'R=TI_N$; $RR'=[I_N\otimes J_T]$; $J_T=j_T j_T'$ is $T\times T$ matrix of ones.

Bayesian Panel Data Model

$$b=(X'HX)^{-1}X'HY \tag{1557}$$

$$a = \overline{Y} - b\overline{X} \tag{1558}$$

Matrix H of dimension $NT \times NT$ is operator for agent specific means as:

$$H_T = I_{NT} - \left(\frac{1}{T}\right) RR' = I_{NT} - \left(\frac{1}{T}\right)([I_N \otimes J_T]) \tag{1559}$$

$$H = \begin{bmatrix} H_T & . & . & 0 & 0 \\ 0 & H_T & . & 0 & 0 \\ . & 0 & . & 0 & 0 \\ . & . & . & H_T & 0 \\ 0 & . & . & 0 & H_T \end{bmatrix}; H_T = I_T - \left(\frac{1}{T}\right) J_T; H_T y = \begin{bmatrix} y_1 - \overline{y} \\ y_2 - \overline{y} \\ . \\ . \\ y_N - \overline{y} \end{bmatrix} \tag{1560}$$

Bayesian Panel Data Model: Differenced Data

$$H_T = D_T' \left(D_T' D_T\right)^{-1} D_T \tag{1561}$$

$$D_T y = \begin{bmatrix} 1 & -1 & . & . & 0 & 0 \\ 0 & 1 & -1 & . & 0 & 0 \\ . & & 0 & . & 0 & 0 \\ . & & . & . & -1 & 0 \\ 0 & & . & . & 1 & -1 \end{bmatrix} \begin{bmatrix} y_T \\ y_{T-1} \\ . \\ . \\ y_1 \end{bmatrix} = \begin{bmatrix} y_T - y_{T-1} \\ y_{T-1} - y_{T-2} \\ . \\ . \\ y_2 - y_1 \end{bmatrix} \tag{1562}$$

$$b = \left(X' D_T' \left(D_T' D_T\right)^{-1} D_T X\right)^{-1} X' D_T' \left(D_T' D_T\right)^{-1} D_T Y \tag{1563}$$

Regression in first differences:

$$\underset{NT\times 1}{D_T y_i} = \underset{NT\times(N+K)}{D_T x_i} \underset{(N+K)\times 1}{\beta} + \alpha_i D_T j_T + \underset{NT\times 1}{D_T e_i} \tag{1564}$$

Bayesian Regression: GAMMA Distribution Gamma family has kernel $p(y) = y^{\alpha-1} e^{-\beta y}$ $y > 0$; $\alpha, \beta > 0$

$$p(y) = \frac{y^{\alpha-1} e^{-\beta y}}{\Gamma(\alpha) \beta^{-\alpha}};\ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \alpha > 0 \tag{1565}$$

when $\alpha = 1$ $\quad p(y) = \beta e^{-\beta y}$ and when $\beta = 1$ then mean and variance equal 1

$$E(Y) = \frac{\alpha}{\beta};\ V(Y) = \frac{\alpha}{\beta^2} \tag{1566}$$

One parameter subfamily of gamma distribution

$$p(y) = \frac{y^{\frac{v}{2-1}} e^{-\frac{y}{2}}}{\Gamma\left(\frac{v}{2}\right) 2^{\frac{v}{2}}}; y > 0; v > 0; \alpha = \frac{v}{2}; \beta = \frac{1}{2}; v > 0 \tag{1567}$$

Bayesian Regression: BETA Distributions Beta distribution depends on two parameters

$$p\left(x/y,\beta\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}x^{\alpha-1}\left(1-x\right)^{\beta-1};\ \alpha,\beta>0;\ 0\le x\le 1$$

$$\Gamma\left(\alpha\right) = \left(\alpha-1\right)! \quad \Gamma\left(1\right)=\left(1-1\right)!=0!=1 \tag{1568}$$

$$\int_0^{\infty} x^{\alpha-1}\left(1-x\right)^{\beta-1}dx = \frac{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta\right)} \tag{1569}$$

The mean and variances of densities

$$E\left(X\right)=\frac{\alpha}{\alpha+\beta};\ V\left(Y\right)=\frac{\alpha\beta}{\left(\alpha+\beta\right)^2\left(\alpha+\beta+1\right)} \tag{1570}$$

Distribution symmetrical around $X=\frac{1}{2}$ if $\alpha=\beta$ and uniform if $\alpha=\beta=1$.

Bayesian Regression: DIRICHIET Distribution Beta distribution depends on two parameters

$$f_p\left(p\right) = \frac{\Gamma\left(\alpha\right)}{\Gamma\left(\alpha_0\right)\Gamma\left(\alpha_1\right)...\Gamma\left(\alpha_L\right)}p_0^{\alpha_0-1}\times ..\times p_L^{\alpha_L-1};\ \alpha,\beta>0;\ 0\le x\le 1$$

$$\left\{p_i\right\} \geq 0;\sum_{i=1}^{L}p_i=1;'\left\{\alpha_i\right\}\geq 0;\sum_{i=1}^{L}\alpha_i=\alpha; \tag{1571}$$

The mean,variance and covariance of Dirichiet distribution

$$E\left(P_j\right)=\frac{\alpha_j}{\alpha};\ V\left(P_j\right)=\frac{\alpha_j\left(\alpha-\alpha_j\right)}{\alpha^2\left(\alpha+1\right)}\ ;C\left(P_iP_j\right)=\frac{\alpha_i\alpha_j}{\alpha^2\left(\alpha+1\right)} \tag{1572}$$

Dirichet is natural conjugate prior family fo rmultinomial distribution of the form

$$y_p\left(y/p\right)=\frac{n!}{y_0!y_1!.....y_L!}p_0^{y_0}p_1^{y_1}\times ..\times p_L^{y_L} \tag{1573}$$

References

[1] http://www.estima.com/cgi-bin/procbrowser.cgi; http://www.bayesian.org/;www.dynare.org; dynare software runs with MATLAB and is used of Bayesian Computations.

[2] Anscombe F. J.(1961) Bayesian Statistics, The American Statistician, 15, 1 , 21-24

[3] Bayes, T. (1763), "An Essay Towards Solving a Problem in the Doctrine of Chances," Phil. Trans. Royal Soc. Of London, 53, 370-418, reprinted in Press (1989).

[4] Bańbura M, D. Giannone and L. Reichlin (2010) Large Bayesian vector auto regressions, Journal of Applied Econometrics 25, 1, 71–92

[5] Bauwens L , M. Lubrano and J. F. Richard (1999) Bayesian Inference in Dynamic Econometric Models, Oxford University Press.

[6] Berger, J., (1990) Robust Bayesian Analysis: Sensitivity to the prior. Journal of Statistical Planningand Inference 25, 303-328.

[7] Bhattarai K. and S. Mallick (2013) Impact of China's currency valuation and labour cost on the US in a trade and exchange rate model. North American Journal of Economicsand Finance, 25:40-59

[8] Canova F, M. Ciccarelli (2004) Forecasting and turning point predictions in a Bayesian panel VAR model, Journal of Econometrics, 120, 2, 327-359

[9] Conley T. G., C. B. Hansen, R. E. McCulloch and P. E. Rossi (2008) A semi-parametric Bayesian approach to the instrumental variable problem , Journal of Econometrics, 144, 1, 276-305

[10] Carriero A, G. Kapetanios and M. Marcellino (2011) Forecasting large datasets with Bayesian reduced rank multivariate models, Journal of Applied Econometrics 26, 5, 735–761

[11] Chamberlain G.(2000) Econometrics and decision theory,Journal of Econometrics 95 255-283

[12] Chetty V. K. and U. Sankar (1969) Bayesian Estimation of the CES Production Function, Review of Economic Studies 36 (3): 289-294

[13] Chib S. (1993) Bayes regression with autoregressive errors:A Gibbs sampling approach, Journal of Econometrics 58, 275-294.

[14] Chib S and E. Greenberg (1994) Bayes inference in regression models with ARMA (p, q) errors, Journal of Econometrics 64 183 -206

[15] Chib S., F. Nardarib and N. Shephardc (2002) Markov chain Monte Carlo methods for stochastic volatility models, Journal of Econometrics 108, 281 316

[16] Chib S., E. Greenberg, R. Winkelmann (1998) Posterior simulation and Bayes factors in panel count data models, Journal of Econometrics 86 33-54

[17] Doan, T., Litterman, R., Sims, C. (1984) Forecasting and conditional projections using realist prior distributions. Econometric Review 3, 1–100.

[18] Dorfman, J.H., (1995) A numerical Bayesian test for cointegration of AR processes. Journal of Econometrics,66, 289–324

[19] Edge R. M., M. T. Kiley and J. P. Laforte (2010) A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model, Journal of Applied Econometrics 25, 4, 720–754

[20] ESTIMA (2004) RATS User Guide, Version 6,Evaston IL 60201 USA.

[21] Fu Q (2000) Bayesian Multivariate Time Series Models for Forecasting European Macroeconomic Series, PhD Dissertation submitted to the University of Hull.

[22] Geweke J. and M. Keane (2000) An empirical analysis of earnings dynamics among men in the PSID: 1968-1989, Journal of Econometrics 96 293-356.

[23] Geweke J (2001) Bayesian econometrics and forecasting, Journal of Econometrics, 100, 1, 11-15

[24] Geweke J. (1996) Bayesian reduced rank regression in econometrics , Journal of Econometrics, 75, 1, 121-146

[25] Geweke J (2005) Contemporary Bayesian Econometrics and Statistics, John Wiley and Sons

[26] Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration. Econometrica ,57, 1317–1339.

[27] George E.I, D, Sun and S Ni (2008) Bayesian stochastic search for VAR model restrictions, Journal of Econometrics 142, 553–580

[28] Harrison P. J. and C. F. Stevens (1976) Bayesian Forecasting, Journal of the Royal Statistical Society. Series B (Methodological), 38, 3, 205-247

[29] Heckman J. J. and T. E. Macurdy (1980) A Life Cycle Model of Female Labour Supply, The Review of Economic Studies, 47,1, 47-74.

[30] Heckelei T and R. C. Mittelhammer (2003) Bayesian bootstrapmultivariate regression, Journal of Econometrics, 112,241– 264

[31] Judge G.G., W.E. Griffiths, R. C. Hill, H. Lutkepohl and T.C. Lee(1990) Theory and Practice of Econometrics, Wiley.

[32] Kleibergen F. R. Paap (2002) Priors, posteriors and bayes factors for a Bayesian analysis of cointegration, Journal of Econometrics, 111, 2, 223-249

[33] Kleibergen F. and E. Zivot (2003) Bayesian and classical approaches to instrumental variable regression, Journal of Econometrics 114, 29 − 72

[34] Koop, G., 1992. Aggregate shocks and macroeconomic fluctuations: a Bayesian approach. Journal of Applied Econometrics 7, 395–411.

[35] Koop G (2003) Bayesian Econometrics, John Wiley and Sons

[36] Koop G., H. K. Van Dijk (2000) Testing for integration using evolving trend and seasonals models: A Bayesian approach, Journal of Econometrics 97 261-291

[37] Levine P, J. Pearlman, G. Perendia and B. Yang (2012) Endogenous Persistence in an estimated DSGE model under imperfect information, EconomicJournal,122,1287–1312

[38] Lancaster T (2004) An Introduction to Modern Bayesian Econometrics, Blackwell

[39] Lancaster T. (1997) Exact Structural Inference in Optimal Job-Search Models, Journal of Business & Economic Statistics, 15, 2,165-179

[40] Li Y, T. Zeng and J. Yu (2014) A new approach to Bayesian hypothesis testing , Journal of Econometrics, 178, Part 3, 602-612

[41] Litterman, R.B., (1986) Forecasting with Bayesian Vector Autoregressions five years of experience. Journal of Business and Economic Statistics 4, 25–38

[42] Musalem A, E. T. Bradlow and J. S. Raju (2009) Bayesian estimation of random-coefficients choice models using aggregate data, Journal of Applied Econometrics, 24, 3, 490–516

[43] Norets A. and X. Tang (2013) Semiparametric Inference in Dynamic Binary Choice Models Review of Economic Studies, online December 15, 2013

[44] Norets A and J. Pelenis (2012) Bayesian modeling of joint and conditional distributions Journal of Econometrics,168, 2, 332-346

[45] Pesaran, M.H., Smith, R. (1996). Estimating long run relationships from dynamic heterogeneous panels,Journal of Econometrics 68, 79–113.

[46] Phillips, P.C.B., Ploberger, W., 1996. An asymptotic theory of Bayesian inference for time series.Econometrica 64, 381–412.

[47] Pratt, J., 1965. Bayesian interpretation of standard inference statements. Journal of the RoyalStatistical Society Series B 27, 169-192.

[48] Rossi P.E. and G. M. Allenby (2003) Bayesian Statistics and Marketing, Marketing Science, 22, 3, 304-328

[49] Rust J. (1996) Non-linear dynamic programming in economics, in Amman, Kendrick and Rust ed. Handbook of Computational Economics, Northholland, Elsevier, Oxford.

[50] Sala L (2014) DSGE Models In The Frequency Domain, Journal of Applied Econometrics, online: 23 JAN 2014

[51] Sims, C., Zha, T., 1998. Bayesian methods for dynamic multivariate models. International Economic Review 39, 949–979.

[52] Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, Journal of the Royal Statistical Society B 55, 3–23.

[53] Zellner R (1985) Bayesian Econometrics, Econometrica, 53:2:253-269.

[54] Schotman P and H.K. van Dijk (1991) A Bayesian analysis of the unit root in real exchange rates,Journal of Econometrics 49 195-238.

[55] Sims C. A. and H. Uhlig (1991) Understanding Unit Rooters: A Helicopter Tour, Econometrica, 59, 6,1591-1599

[56] Zellner A. (1988) Bayesian analysis in econometrics , Journal of Econometrics, 37, 1, 27-50

[57] Zellner A. (1985) Bayesian Econometrics, Econometrica, 53, 2, 253-270.

10.4.4 Tutorial 6: Bayesian Modelling

Q1. Write short notes on any five of the following issues relating to the Bayesian modelling and analysis.

a. Difference between classical and Bayesian assumptions on parameters and errors in a linear regression.

b. Bayesian rule where $p(A, B)$ denotes the joint probability of occurring events A and B together, $p(B)$ is the marginal probability of B without any respect to occurrence of A. Similarly $p(A)$ is the marginal probability of A without any respect to occurrence of B.

c. Bayesian prior and posterior density functions on unknown parameters β and τ for the likelihood function such as: $\log l(\beta,\tau) = \frac{1}{2}\left[n\log\tau - \tau(\beta-b)X'X(\beta-b) - \tau e'e\right]$

d. Estimates of the mean and variance of β and τ in the above Bayesian linear regression model.

e. Bayesian panel data model of the form $y_{i,t} = x_{i,t}\beta + \alpha_i + e_{i,t}$ with priors $e_{i,t}/x_{i,t},\beta,\alpha_i,\tau \sim N(0,\tau)$.

f. Estimation procedure in a Bayesian VAR model of the form: $\underset{T\times 1}{Y_j} = \underset{T\times K}{X_j}\ \underset{K\times 1}{\beta_j} + \underset{T\times 1}{u_j}$ where $u_j \sim N\left(0,\sigma_{i,j}^2 I_T\right)$.

g. Specification and estimation procedure of the Bayesian stochastic volatility models (Bayesian ARCH/GARCH).

h. MCMC algorithm. Koop's MATLAB routines.

11 L10: Generalised Method of Moments (GMM)

Generalised Method of Moments (GMM) is the most general estimation technique to estimate the unknown parameters in an econometric model. Originally developed by Hansen (1982), though tts antecedents include methods of moments by Pearson (1893, 1894 and 1895) and instrumental variables (Wright (1925), Sargan (1958). It has become very popular in recent years Tauchen (1986) Newey (1988) Hamilton(1994)Andersen and Sorensen (1996), Hansen, Heaton and Yaron (1996) Gallant and Tauchen (1996) Smith(1997), Arrelano-Bond (1991), Blundell-Bond (1998), Andrews DWK (1999), Eviews (Chapter 20), STATA (pages 571-531), Stock -Wright-Yogo (2002), Wiendmeijer (2000) Hall (2003) Newey and Smith (2004), Greene (2008), Tripathi (2011).

Literature on GMM

- Sargan(1958), Hansen (1982), Hansen and Singleton(1982), Tauchen (1986), Newey (1988), Arellano and Bond (1991), Hansen and Scheinkman, (1995), Andersen and Sorensen (1996), Hansen, Heaton and Yaron (1996), Gallant and Tauchen (1996), Smith (1997), Kitamura and Phillips (1997), Kitamura and Stutzer(1997), Blundell and Bond (1998), Imbens, Spady and Johnson (1998), Hall (2003), Newey and Smith (2004), Windmeijer (2005), Bera and Bilias (2002), Wright (2003)

- Smith (1997), Smith and Ramalho (2013) , Smith (2011), Smith and Parente (2011), Richard , Taylor and Castro (2009), Smith (2005), Smith and Guggenberger (2005), Smith and Newey, (2004), Imbens (1997), Antoine and Renault (2009), Heckman and Todd (2009), Li and Muller (2009), Beaulieu, Browning, Ejrnaes and Alvarez (2010), Bun and Windmeijer (2010), Wright (2010), Gørgens and Würtz(2012), Creel and Kristensen (2012), Dufour and Khalaf (2013), Doraszelski and Jaumandreu (2013), Leslie and Sorensen (2014)

- Quintos (1998), Koenker and Machado (1999), Ronchetti and Trojani (2001), Bera and Bilias (2002), Gagliardini, Trojani and Urga (2005), Kleibergen (2007), Jondeau and Bihan (2008), Tripathi (2011), Kuersteiner (2012).

- First introduced by Hansen (1982). Encompasses many estimators in econometrics.

- Convenient method of estimating non-linear dynamic models withouth complete knowledge of the probability distributions of the data.
- Applications in time series, cross section and panel data and wide range of fields in economics.
- Its antecedents include methods of moments by Pearson (1893, 1894 and 1895) and instrumental variables (Wright (1925), Sargan (1958).
- Consider a density of variable v with θ parameters as $f(v,\theta)$ where v is $q\times 1$ vector and θ $p\times 1$ vector of parameters. Traditions econometrics is application for $q=p$ case but the GMM can apply even if $q>p$.
- In $Q_n(\theta)=g_n(\theta)' W_n g_n(\theta)$ for a given weight matrix W_n the GMM estimator is defined as $\theta_n=\arg\min_{\theta\in\Theta} Q_n(\theta)$ where Θ denotes the parameter space.

Introduction

- Consider a problem of estimating population mean μ of variable y_i on the basis of $i=1,2,.....,N$ sample obsrvations. Its moment condition is expressed as

$$E\{y_i-\mu\}=0 \tag{1574}$$

Its sample equivalent form is

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu)=0 \tag{1575}$$

Therefore

$$\widehat{\mu}=\frac{1}{N}\sum_{i=1}^{N}y_i \tag{1576}$$

- This is just the sample average estimated using the moment condition.

11.1 GMM for a Linear Regression

For a linear regression $y_i=x_i'\beta+\varepsilon_i$ with instrument z_i the moment condition is expressed as

$$E\{\varepsilon_i z_i\}=E\left\{\left(y_i-x_i'\beta\right)z_i\right\}=0 \tag{1577}$$

Main problem of GMM estimation is to find an optimal weighting (covariance) matrix W_n

$$\widetilde{\beta}=\left(X'ZW_nZ'X\right)^{-1}XZW_nZ'y \tag{1578}$$

when ε_i is i.i.d. the optimal weight is

$$W^{opt}=\left[E\left\{\varepsilon_i^2 z_i z_i'\right\}^{-1}\right]=\left[\frac{1}{N}\left\{\sum_{i=1}^{N}\widehat{\varepsilon}_i^2 z_i z_i'\right\}\right]^{-1}=\left[\frac{1}{N}\sum_{i=1}^{N}z_i z_i'\right]^{-1} \tag{1579}$$

Covariance matrix

$$V = \left(DW^{opt}D'\right)^{-1} = \left[\frac{\partial E\left\{\left(y_i - x_i^{'}\beta\right)z_i\right\}}{\partial\beta}\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N} z_i z_i^{'}\right]$$
$$\left[\left(\frac{\partial E\left\{\left(y_i - x_i^{'}\beta\right)z_i\right\}}{\partial\beta}\right)^{'}\right]^{-1} \tag{1580}$$

$$V\left(\widetilde{\beta}\right) = \left(DW^{opt}D'\right)^{-1} = \left[\sum_{i=1}^{N} x_i z_i^{'}\right]^{-1}\sum_{i=1}^{N}\widehat{\varepsilon}_i^2 z_i z_i^{'}\left[\sum_{i=1}^{N} x_i z_i^{'}\right]^{-1} \tag{1581}$$

Orthogonality conditions

$$E\left\{\underset{r\times 1}{h}\left(\theta_0, w_t\right)\right\} = 0 \tag{1582}$$

where w_t are strictly stationary variables observed at date t and θ_0 are true value of $a \times 1$ vector of unknown parameters and $h(.)$ is differentiable r-dimentional vector valued functions with $r. > a$. GMM estimate θ_0 is the value of θ that minimise

$$\underset{1\times r}{g\left(\theta, Y_t\right)'}\underset{r\times r}{\widehat{S}_T^{-1}}\underset{r\times 1}{g\left(\theta, Y_t\right)} = 0 \tag{1583}$$

$$\underset{1\times r}{g\left(\theta, Y_t\right)} = \frac{1}{T}\sum_{t}^{T}\underset{r\times 1}{h\left(\theta, w_t\right)} \tag{1584}$$

Orthogonality conditions

where S_T is the estimate of

$$S = \underset{t\longrightarrow\infty}{lim}\left(\frac{1}{T}\right)\sum_{t}^{T}\sum_{v=-\infty}^{\infty}E\left[h\left(\theta_0, w_t\right)\right]\left[h\left(\theta_0, \underset{1\times r}{w_{t-v}}\right)\right] \tag{1585}$$

GMM encompases many estimators including the OLS, MLE, instrumental variable, two stage least square, non-linear simultaneous equation, dynamic rational expectation and non-staionary data models.

Ruud (2000) and Hamilton (1994) have good chapters in GMM.

Orthogonality conditions: An Example

Consider a standard regression model $Y = \beta X + e$ where X is $K \times 1$ explanatory variables

$$E\left(X_t u\right) = 0 \tag{1586}$$

$$E\left(X_t\left(Y - X'\beta\right)\right) = 0 \tag{1587}$$

This is equivalent to K orthogonality conditions of the GMM with $w_t = \{Y_t, X_t\}$ and $\theta = \beta$

$$h\left(\theta_0, w_t\right) = X_t\left(Y - X'\beta\right) \tag{1588}$$

number of orthogonality conditions equal the number of parameters $r = a = k$.

11.1.1 OLS is a special case of GMM

Thus the standard regression model is just a special case of GMM specification

$$0 = g(\theta, Y_t) = \frac{1}{T}\sum_{t}^{T}\left(X_t\left(Y_t - X_t'\beta\right)\right) \tag{1589}$$

Rearranging this

$$\sum_{t}^{T} X_t Y_t = \left\{\sum_{t}^{T} X_t X_t'\right\}\widehat{\beta}_T \tag{1590}$$

$$\widehat{\beta}_T = \left\{\sum_{t}^{T} X_t X_t'\right\}^{-1}\sum_{t}^{T} X_t Y_t \tag{1591}$$

Hence the OLS is the special case of the GMM estimation.

11.1.2 MLE is a special case of GMM

Similarly the maximul likelihood function implies

$$L(\theta) = \sum_{t}^{T}\log f\left(Y_t / Y_{t-1}; \theta\right) \tag{1592}$$

First order conditions

$$\sum_{t}^{T}\frac{\partial \log f\left(Y_t / Y_{t-1}; \theta\right)}{\partial \theta} = 0 \tag{1593}$$

$$0 = \frac{1}{T}\sum_{t}^{T} h(\theta, y_t) \tag{1594}$$

Thus the orthogonality condition to estimate θ are same in both GMM and MLE.
Instrumental Variable Estimation:An Example

$$y = X\beta_0 + u \tag{1595}$$

$$Q_n(\beta) = \left[n^{-1}u(\beta)' Z\right] W_n \left[n^{-1}Z' u(\beta)\right] \tag{1596}$$

$$Q_n(\beta) = \left[n^{-1}(y - X\beta)' Z\right] W_n \left[n^{-1}Z'(y - X\beta)\right] \tag{1597}$$

First order conditions for optimisation imply:

$$-2X^{'}ZW_nZ'y + 2XZW_nZ'X\widetilde{\beta} = 0 \tag{1598}$$

$$X^{'}ZW_nZ'y = XZW_nZ'X\widetilde{\beta} \tag{1599}$$

$$\widetilde{\beta} = \left(Z'X\right)^{-1} Z'y \tag{1600}$$

Moment estimates

$$\widetilde{\beta} = \left(X^{'}ZW_nZ'X\right)^{-1} XZW_nZ'y \tag{1601}$$

$$n^{\frac{1}{2}}\left(\beta - \beta_0\right) = \left(n^{-1}X^{'}ZW_nn^{-1}Z'X\right)^{-1} n^{-1}X^{'}ZW_nn^{-\frac{1}{2}}Z'u \tag{1602}$$

11.2 Basics of GMM

population moments

$$E\left[f\left(x_t, \beta_0\right)\right] = 0 \tag{1603}$$

Sample moments

$$g_n\left(\beta\right) = n^{-1}\sum_{t=1}^{n} f\left(x_t, \beta_0\right) \tag{1604}$$

objective quadratic function

$$Q_n\left(\beta\right) = g_n\left(\beta\right)' W_n g_n\left(\beta\right) \tag{1605}$$

first order conditions

$$G_n\left(\widetilde{\beta}\right)' W_n g_n\left(\widetilde{\beta}\right) = 0 \tag{1606}$$

$G_n\left[\beta\right]_{ij} = \frac{\partial g_{ni}(\beta)}{\partial \beta_j}$ and W_n is positive definite weigth matrix that converges to W.
Asymptotics

$$n^{-1}\left(\widetilde{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, V_G\right) \tag{1607}$$

where $V_G = \left[G_0^{'}WG_0\right]^{-1} G_0^{'}WS_wWG_0\left[G_0^{'}WG_0\right]^{-1}$ and $G_0 = E\left[\frac{\partial f(x_t,\beta_0)}{\partial \beta}\right] = 0$

11.2.1 GMM and Euler Equation Model:A CAPM Example

Hansen and Singleton (1982) example is widely cited in the GMM literature.
It is consumer optimisation problem stated as:

$$max\ E_0\left[\sum_{t=0}^{\infty} \delta^t U\left(C_t\right)\right] \tag{1608}$$

subject to

$$C_t + P_tQ_t \leq R_tQ_{t-1} + W_t \tag{1609}$$

Lagrangian

$$L_t = E_0 \left[\sum_{t=0}^{\infty} \delta^t U(C_t) \right] + \sum_{t=0}^{\infty} \lambda_t \left[C_t + P_t Q_t - R_t Q_{t-1} - W_t \right] \tag{1610}$$

The first order condition for optimisation implies

$$P_t U'(C_t) = \beta E_t \left[R_{t+1} U'(C_{t+1}) \right] \tag{1611}$$

$$E_t \left[\delta \frac{R_{t+1}}{P_t} \frac{U'(C_{t+1})}{U'(C_t)} - 1 \right] = 0 \tag{1612}$$

Specify the utility function as with $\gamma < 1$

$$U(C_t) = \frac{C_t^{\gamma}}{\gamma} \tag{1613}$$

$$E_t \left[\delta \frac{R_{t+1}}{P_t} \left[\frac{C_{t+1}}{C_t} \right]^{\alpha} - 1 \right] = 0 \tag{1614}$$

$\alpha = \gamma - 1$

How to estimate α and δ in this model? Maximum likelihood is computationally burdensome and generates biased results. In contrast, the GMM can be applied using the moment conditions as:

$$E \left[\delta \frac{R_{t+1}}{P_t} \left[\frac{C_{t+1}}{C_t} \right]^{\alpha} - 1 \right] = E \left\{ E_t \left[\delta \frac{R_{t+1}}{P_t} \left[\frac{C_{t+1}}{C_t} \right]^{\alpha} - 1 \right] \right\} = 0 \tag{1615}$$

With information set Ψ use instruments $z_t \in \Psi$ and $y_{t+1} \notin \Psi$. With $E_t [y_{t+1}] = 0$

$$E_t [y_{t+1} z_t] = \{ E_t [y_{t+1}] \} z_t = 0 \tag{1616}$$

$$E_t [u_{t+1}(\alpha, \delta) z_t] = 0 \tag{1617}$$

where $u_{t+1}(\alpha, \delta) = \left[\delta \frac{R_{t+1}}{P_t} \left[\frac{C_{t+1}}{C_t} \right]^{\alpha} - 1 \right]$ and z_t is $r \times 1$ vector of instruments. Instruments z_t may include past values of variables such as $C_{t-i}, P_{t-i}, R_{t-i}$ for $i \geq 0$.

Notice that it is not necessary to explicitly specify the data generating process like in OLS or LME here as $u_{t+1}(\alpha, \delta)$ is just conditional expectation of data with conditional expectation zero.

GMM is computationally more convenient and avoids bias due to misspecification compared to the MLE. (see Smith (1997) for efficiency comparison between GMM and MLE).

Arrelano-Bond estimator

First consider a dynamic panel data model that takes the form:

$$y_{it} = \alpha_i + \rho y_{it-1} + x_{i,t} \beta + u_{i,t} \tag{1618}$$

difference of this equation is

$$\Delta y_{it} = \alpha_i + \rho \Delta y_{it-1} + \Delta x_{i,t} \beta + \Delta u_{i,t} \tag{1619}$$

More compactly write it as:

$$\Delta y = \Delta R\pi + \Delta u \tag{1620}$$

with orthogonality condition

$$E\left(\Delta y_{it-1} u_{i,t}\right) = 0 \tag{1621}$$

$$\pi_{GMM} = \left(\Delta R' Z (Z'\Omega Z)^{-1}) Z' \Delta R\right)^{-1} \Delta R' Z (Z'\Omega Z)^{-1}) Z' \Delta y \tag{1622}$$

where Z is instrument for ΔR. Use y_{it-2} or the higher lags of y_{it} as the instrument set.

Blundell and Bond (1998) derived conditions to use additional sets of moment conditions to improve the small sample perfomance of Arrelano-Bond estimator

$$E\left(\Delta y_{it-1}\left(\alpha_i + u_{i,t}\right)\right) = 0 \tag{1623}$$

$$E\left(Z_{sys}^T P_i\right) = 0 \tag{1624}$$

$$Z_{sys}^T = \begin{bmatrix} z_{it-1} & 0 & 0 & 0 & 0 \\ 0 & \Delta y_{i2} & 0 & 0 & 0 \\ 0 & 0 & \Delta y_{i3} & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & . \end{bmatrix} \text{ and } P_i = \begin{bmatrix} \Delta u_i \\ u_{i,3} \\ u_{i,4} \\ u_{i,5} \\ . \end{bmatrix}$$

with these instruments they estimate the system GMM

$$\pi_{GMM}^{Sys} = \left(\Delta R' Z (Z'\Omega Z)^{-1}) Z' \Delta R\right)^{-1} \Delta R' Z (Z'\Omega Z)^{-1}) Z' \Delta y \tag{1625}$$

GMM Estimation

In Eviews open macro08_uk.csv and WK1 file. Object/New object/equation/ specify dependent, independent variables and instruments.

Dependent Variable: CONS_HH
Method: Generalized Method of Moments
Date: 05/10/10 Time: 20:54
Sample (adjusted): 1966Q1 2006Q1
Included observations: 161 after adjustments
Linear estimation with 2 weight updates
Estimation weighting matrix: HAC (Bartlett kernel, Newey-West fixed
bandwidth = 5.0000)
Standard errors & covariance computed using estimation weighting matrix
Instrument specification: EXPORTS FINCONS TBILLS
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M4	0.023318	0.002617	8.909216	0.0000
GDP_MP	0.494808	0.044538	11.10982	0.0000
PUB_CONS	0.615099	0.427344	1.439352	0.1520

R-squared	0.993050	Mean dependent var	93764.26
Adjusted R-squared	0.992962	S.D. dependent var	30303.00
S.E. of regression	2542.113	Sum squared resid	1.02E+09
Durbin-Watson stat	1.391309	J-statistic	7.204895
Instrument rank	4	Prob(J-statistic)	0.007271

Orthogonality Test

ıstrument Orthogonality C-test Test
quation: CONSUM
pecification: CONS_HH M4 GDP_MP PUB_CONS
ıstrument specification: EXPORTS FINCONS TBILLS
est instruments: EXPORTS

	Value	df	Probability
ifference in J-stats	7.204895	1	0.0073

-statistic summary:

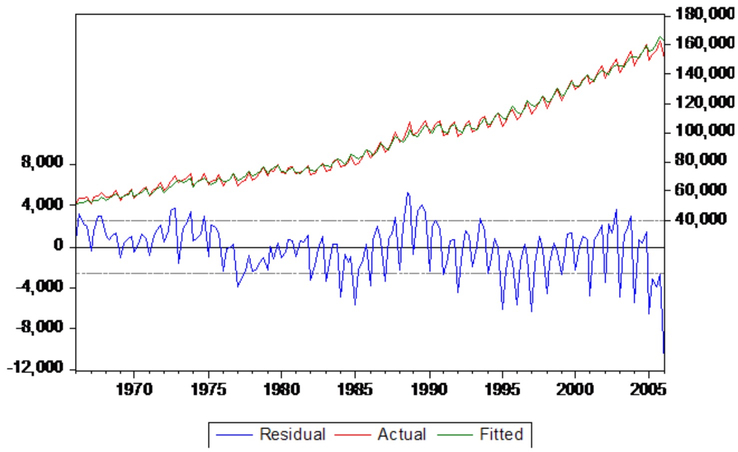
	Value
estricted J-statistic	7.204895
nrestricted J-statistic	0.000000

nrestricted Test Equation:
ependent Variable: CONS_HH
lethod: Generalized Method of Moments
ate: 05/10/10 Time: 21:02
ample (adjusted): 1966Q1 2006Q1
ıcluded observations: 161 after adjustments
ixed weighting matrix for test evaluation
tandard errors & covariance computed using estimation weighting matrix
ıstrument specification: C FINCONS TBILLS

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M4	0.022023	0.002637	8.351059	0.0000
GDP_MP	0.488217	0.044189	11.04832	0.0000
PUB_CONS	0.709165	0.424792	1.669439	0.0970

-squared	0.993296	Mean dependent var	93764.26
djusted R-squared	0.993211	S.D. dependent var	30303.00
.E. of regression	2496.764	Sum squared resid	9.85E+08
urbin-Watson stat	1.487252	J-statistic	0.000000
ıstrument rank	3		

11.2.2 Prediction and Residual



Hansen's GMM estimation (see Hansen.prg)
Arellano-Bond estimator;The instruments need to be constructed (see ARELLANO.PRG)
Nonlinear instrumental variable method (GIV.prg)
Instrumental variable model of logwage rate by Wooldridge (INSTRUMENT.PRG
. global xb "{b1}*gear_ratio +{b2}*length +{b3}*headroom + {b0}"
. global phi "normalden($xb)"
. global phi "normal($xb)"
. estimate store ml
estimates store ml
. estimates store gmm

. estimates table ml gmm, b se

Variable	ml	gmm
mpg	-224.35974	-224.35974
	63.486415	63.486415
trunk	126.60486	126.60486
	104.3013	104.3013
headroom	-659.46302	-659.46302
	471.23331	471.23331
_cons	11175.774	11175.774
	2364.5154	2364.5154

legend: b/se

GMM in STATA
Statistics/Endogenous covariates

```
. ivregress 2sls price mpg headroom weight (rep78 = length turn)

Instrumental variables (2SLS) regression               Number of obs =      69
                                                       Wald chi2(4)  =   38.66
                                                       Prob > chi2   =  0.0000
                                                       R-squared     =  0.3589
                                                       Root MSE      =  2314.9
```

price	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
rep78	1517.804	867.8677	1.75	0.080	-183.1851	3218.794
mpg	-83.29217	87.50904	-0.95	0.341	-254.8067	88.2224
headroom	-781.6608	380.0685	-2.06	0.040	-1526.581	-36.74023
weight	2.677978	.6782855	3.95	0.000	1.348563	4.007393
_cons	-3024.737	4459.78	-0.68	0.498	-11765.75	5716.271

```
Instrumented:  rep78
Instruments:   mpg headroom weight length turn

. reg3 (price = mpg trunk headroom)

Three-stage least-squares regression
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
price	74	3	2550.05	0.2423	23.66	0.0000

price	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
price						
mpg	-224.3597	63.48641	-3.53	0.000	-348.7908	-99.92865
trunk	126.6049	104.3013	1.21	0.225	-77.82194	331.0317
headroom	-659.463	471.2333	-1.40	0.162	-1583.063	264.1373
_cons	11175.77	2364.515	4.73	0.000	6541.409	15810.14

```
Endogenous variables:  price
Exogenous variables:   mpg trunk headroom
```

```
. ivregress liml price mpg headroom weight (rep78 = length turn)

Instrumental variables (LIML) regression               Number of obs =      69
                                                       Wald chi2(4)  =   22.82
                                                       Prob > chi2   =  0.0001
                                                       R-squared     =       .
                                                       Root MSE      =  3058.5

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rep78 |   3200.096   2055.613     1.56   0.120     -828.832    7229.025
         mpg |  -148.5115   133.1996    -1.11   0.265    -409.5779    112.5549
    headroom |  -900.8467   516.4899    -1.74   0.081    -1913.148     111.455
      weight |   3.191665   1.036573     3.08   0.002     1.160018    5.223311
       _cons |  -8565.727   8142.278    -1.05   0.293     -24524.3    7392.845
------------------------------------------------------------------------------
Instrumented:  rep78
Instruments:   mpg headroom weight length turn
```

```
. ivregress gmm price mpg headroom weight (rep78 = length turn), igmm vce(bootstrap)
(running ivregress on estimation sample)

Bootstrap replications (50)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50

Instrumental variables (GMM) regression                Number of obs =      69
                                                       Wald chi2(4)  =   22.58
                                                       Prob > chi2   =  0.0002
                                                       R-squared     =  0.2591
GMM weight matrix: Robust                              Root MSE      =  2488.6

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       price |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rep78 |    1879.44   710.0672     2.65   0.008     487.7336    3271.146
         mpg |  -146.6322   79.13221    -1.85   0.064    -301.7285    8.464096
    headroom |  -639.6306   330.3725    -1.94   0.053    -1287.149     7.88748
      weight |   1.842779   .6897574     2.67   0.008     .4908799    3.194679
       _cons |  -1132.796   3582.763    -0.32   0.752    -8154.882     5889.29
------------------------------------------------------------------------------
Instrumented:  rep78
Instruments:   mpg headroom weight length turn
```

```
. ivregress gmm price mpg headroom weight (rep78 = length turn), igmm vce(jackknife)
(running ivregress on estimation sample)

Jackknife replications (69)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
...................

Instrumental variables (GMM) regression                Number of obs =      69
                                                       Replications  =      69
                                                       F(  4,    68) =    4.80
                                                       Prob > F      =  0.0018
GMM weight matrix: Robust                              R-squared     =  0.2591
                                                       Root MSE      =  2488.6

------------------------------------------------------------------------------
             |              Jackknife
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       rep78 |    1879.44   627.5589     2.99   0.004     627.1654    3131.714
         mpg |  -146.6322   87.25377    -1.68   0.097    -320.7444    27.47999
    headroom |  -639.6306   333.6885    -1.92   0.059    -1305.496    26.23439
      weight |   1.842779   .8101387     2.27   0.026     .2261729    3.459386
       _cons |  -1132.796   4049.643    -0.28   0.781    -9213.734    6948.141
------------------------------------------------------------------------------
Instrumented:  rep78
Instruments:   mpg headroom weight length turn
```

References

[1] Andersen TG and BE Sorensen (1996) GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, Journal of Business and Economic Statistics, 44:3:Jul: 328-352.

[2] Antoine, B. and Renault, E. (2009), Efficient GMM with nearly-weak instruments. The Econometrics Journal, 12: S135–S171

[3] Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. Review of Economic Studies 58 (2), 277–297.

[4] Arellano, M., Bover, O., 1995. Another look at the instrumental variable esti-mation of error-components models. Journal of Econometrics 68 (1), 29–51.

[5] Basu P. and K. Bhattarai (2012) Cognitive Skills, Openness and Growth, the Economic Record, 88: 280: March, 18-38

[6] Basu P., K. Bhattarai and Y. Getachew (2013) Capital adjustment cost and bias in income based dynamic panel models with fixed effects, mimio, Universities of Durham and Hull, UK.

[7] Beaulieu M.C., J. Dufour and L. Khalaf (2013) Identification-Robust Estimation and Testing of the Zero-Beta CAPM Review of Economic Studies 80 (3): 892-924

[8] Bera A.K. and Y. Bilias (2002) The MM, ME, ML, EL, EF and GMM approaches to estimation: a synthesis, Journal of Econometrics, 107, 1–2, March, 51-86

[9] Blundell, R., and S. Bond (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics,* 87, 115–143.

[10] Browning M, M. Ejrnaes, and J. Alvarez (2010) Modelling Income Processes with Lots of Heterogeneity Review of Economic Studies 77 (4): 1353-1381

[11] Bun M. J. G. and Windmeijer, F. (2010), The weak instrument problem of the system GMM estimator in dynamic panel data models. The Econometrics Journal, 13: 95–126

[12] Creel, M. and Kristensen, D. (2012), Estimation of dynamic latent variable models using simulated non-parametric moments. The Econometrics Journal, 15: 490–515.

[13] Doraszelski U. and J. Jaumandreu (2013) R&D and Productivity: Estimating Endogenous Productivity Review of Economic Studies 80 (4): 1338-1383

[14] Gagliardini P., F. Trojani and G Urga (2005) Robust GMM tests for structural breaks, Journal of Econometrics, 129, s 1–2, November–December 2005, 139-182

[15] Gallant R A and G Tauchen (1996) Which Moments to Match? Econometric Theory, 12:4:Oct: 657-681.

[16] Gørgens, T. and Würtz, A. (2012), Testing a parametric function against a non-parametric alternative in IV and GMM settings. The Econometrics Journal, 15: 462–489

[17] Greene W. (2008) Econometric Analysis, Sixth Edition,Pearson Education, New York.

[18] Hall A (2003) Generalised Methods of Moments in Baltagi ed.Theoretical Econometrics, Blackwell.

[19] Hall A (1993) Some aspects of generalised method of moemtns estimation, Maddala, Rao and Vinod eds., Handbook of Statistics, vol. 11,pp.393-417.

[20] Hansen L.P. (1982) Large sample properties of generalized method of moment estimators, Econometrica, 50:4:1029-1054.

[21] Hansen L.P. and K J Singleton(1982) Generalized Instrumental Variables Estimation of Non-linear Rational Expectations Models, Econometrica, 50:5:1269-1286.

[22] Hansen, L. P. and Scheinkman, J. A. (1995). 'Back to the future: generating moment implications for continuous-time Markov processes.' Econometrica, 63, 767-804

[23] Hansen L.P. J. Heaton and A Yaron (1996) Finite Sample Properties of Some Alternative GMM Estimators, Journal of Business and Economic Statistics, 14:3:Jul: 262-280.

[24] Hamilton JD (1994) Time Series Analysis, Princeton.

[25] Heckman, . J. J. and Todd, . P. E. (2009), A note on adapting propensity score matching and selection models to choice based samples. The Econometrics Journal, 12: S230–S234

[26] Hendrry David (1995) Dynamic Econometrics, Oxford University Press.

[27] Imbens G.W. (1997) One-Step Estimators for Over-Identified Generalized Method of Moments Models Review of Economic Studies , 64 (3): 359-383

[28] Imbens, G.W., Spady, R.H., Johnson, P. (1998) Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

[29] Jondeau E and H. L. Bihan (2008) Examining bias in estimators of linear rational expectations models under misspecification, Journal of Econometrics, 143, 2, 375-395

[30] Kitamura, Y., Stutzer, M., 1997. An information theoretic alternate to generalized method of moments

estimation. Econometrica 65, 861–874

[31] Kitamura Y, P. C.B. Phillips (1997) Fully modified IV, GIVE and GMM estimation with possibly non-stationary regressors and instruments Journal of Econometrics, 80, 1, September , 85-123.

[32] Kleibergen F. (2007) Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics, Journal of Econometrics, 139, 1, 181-216

[33] Koenker R. and J. A.F. Machado (1999) GMM inference when the number of moment conditions is large, Journal of Econometrics, 93, 2, 327-344

[34] Kuersteiner G. M. (2012) Kernel-weighted GMM estimators for linear time series models, Journal of Econometrics, 170,2, 399-421

[35] Leslie P. and A Sorensen (2014) Resale and Rent-Seeking: An Application to Ticket Markets Review of Economic Studies 81 (1): 266-300

[36] Li H. and U. K. Muller (2009) Valid Inference in Partially Unstable Generalized Method of Moments Models Review of Economic Studies , 76 (1): 343-365

[37] Newey W K (1988) Asymptotic Equivalnce of Closet Moments and GMM Estimators,Econometric Theory, 4:2:Aug:336-340.

[38] http://www.stata-press.com/data/r11/

[39] Newey W K and R J Smith (2004) Higher order properties of GMM and Eneralised Likelihood Estimators, Econometrica, 72:1:Jan:219-255.

[40] Quintos C. E (1998) Analysis of cointegration vectors using the GMM approach , Journal of Econometrics, 85, 1, 155-188

[41] Ronchetti E and F Trojani (2001) Robust inference with GMM estimators, Journal of Econometrics, 101, 1, 37-69

[42] Ruud P A (2000) An Introduction to Classical Econometric Theory, Oxford.

[43] Roodman, D., 2009. A note on the theme of too many instruments. Oxford Bulletin of Economics and Statistics 71 (1), 135–158.

[44] Sargan, J.D., 1958. The estimation of oconomic relationships using instrmnenial variables.Economctrica 26, 393-415.

[45] Smith R J (1997) Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation, Economic Journal 107:March, 503-519

[46] Tauchen G (1986) Statistical properties of generalized methods of moments estimators of structural parameters obtained from financial marekt data, Journal of Business and Economic Statistics, 4:4:Oct: 397-416.

[47] Tripathi G (2011) Generalized method of moments (GMM) based inference with stratified samples when the aggregate shares are known, Journal of Econometrics, 165, 2, 258-265

[48] Verbeek M. (2012) A Guide to Modern Econometrics, 4th edition, Wiley.

[49] Windmeijer, F.(2005) A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126: 25–51.

[50] White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, New York.

[51] Wright, . J. H. (2010), Testing the adequacy of conventional asymptotics in GMM. The Econometrics Journal, 13: 205–217.

[52] Wright J H (2003) Detecting Lack of Identification in GMM, Econometric Theory, 19:2:Apr: 322-330.

12 Multivariate Analysis: Short Notes

12.1 Objective:

Generating knowledge by processing quantitative and qualitative data using standard statistical techniques; finding patterns in vast amount of information contained in surveys, statistical reports.

Dependent Analysis:

- Multiple regression - Metric Variables
- Time series analysis
- Conjoint analysis - categorical dependent variables
- Discriminant analysis - categorical dependent variables
- Multivariate analysis of variance (MANOVA)
- Canonical correlations
- Structural equation modelling

Interdependent Analysis

- Factor analysis
- Cluster analysis
- Multidimensional scaling
- Correspondence analysis.

Standard approach to Multivariate Modelling

1. defining the research problem - aims and objectives and expected outcomes
2. develop an analytical plan - methods of analysis
3. evaluate the assumption underlying the analysis
4. estimate the multivariate model and assess its overall fit
5. interpret variates
6. validate the multivariate model
7. decision flow chart

12.2 Data Structure:

- Univariate representation - histograms, frequency- stem and leaf
- Bivariate - box plots
- Multivariate: graphical representation
- Data transformation (logs, differences, multiplications, exponentiation)

12.3 Underlying assumption

- Normality - mean, skewness and Kurtosis tests
- Homoscedasticity
- Linearty
- No autocorrelation in errors

Problems of missing observations or outliers

- Extent and randomness of missing data
- Missing at random
- Missing completely at random
  Imputation methods
- Problems of outliers
- Univariate detection
- Bivariate detection
- Multivariate detection.

12.4 Dependent Analysis:

One or multiples of dependent variable explained by a set of independent variables.

- Multiple regression - Metric Variables
- Correlation analysis
- Regression X-variables cause Y

$$Y = f\left(X_1, X_2, ..., X_n\right)$$

magnitude of this relation given by regression coefficients and their statistical significance (T, F, R-square tests)

- Best, Linear and Unbiased estimators of coefficients
- Assumption: No multicollinearity, Homoscedaticity, no autocorrelation
  Consequences of violation of these assumptions
- Detection of collinearity, hetroscedasticity and autocorrelation and remedial measures to solve problems.
  Type I and II errors
  Examples

12.5 Discriminant analysis - categorical dependent variables

Statistical technique to compare the means of a set of independent variables for two or more groups. Males differ from female, North- South, skilled vs unskilled.

- Categorical dependent variable
- Logistic regression
- Frequency distributions for two/multiple of groups.
- Steps for modelling discriminant analysis
- Sort observations by groups and check the differences in group means
- Determine independent variables and sample size
- Assumptions - normality, linearity, no multicollinearity, equal dispersions
- Estimation of the characteristic function - simultaneous or stepwise

- Check the significance of the discriminant function
- Assess prediction accuracy
- Interpret results - discriminant weights, loadings, partial fractions, split samples and cross validation
  Examples

12.6 Conjoint analysis - categorical dependent variables

Evaluating how customers develop preferences for any type of object
Utility based on

- Attributes of the product (colour, taste, content)
- Overall utility by factors, level and treatment elements (solid, liquid; Brand)
- Conjoint task

Determining the total worth of the product
Total worth of the product i ,j $n_{i,j}$
= Part worth of level of factor 1+ Part worth of level of level j factor2 + Part worth of level of level j factor m
Utility, factors, level, stimuli

- Goodness of fit measures: Pearson, Kendal's tau for estimation and validation samples; Part worth of estimates, rescaling and reversals, validation of results

12.7 Multivariate analysis of variance (MANOVA)

- Assessing group differences
- t-test on differences of mean for two variables
- ANOVA
  Total variance
  Within group variances
  Between group variance
- F test; level of significance and critical values
- Multivariate analyses of variance (MANOVA)
  Hotellings T^2 test

$$T^2 = \frac{p\,(N_1 + N_2 - 2)}{N_1 + N_2 - p - 1} \times F_{crit}$$

Multivariate analysis of covariance

In a multivariate analysis

$$Y = X\beta + \epsilon$$

leads to multivariate hypotheses of the form

$$C\beta A = 0$$

where β is a matrix of parameters, C specifies constraints on the design matrix X for a particular hypothesis, and A provides a transformation of Y. A is often the identity matrix. An estimate of is provided by

$$B = (X'X)^{-1}X'Y$$

W = Wilks' lambda L = Lawley-Hotelling trace ; P = Pillai's trace R = Roy's largest root.

The inclusion of weights, if specified, enters the formulas in a manner similar to that shown Methods and formulas in STATA Manual.

All four tests are admissible, unbiased and invariant. Asymptotically, Wilks's lambda, Pillai's trace, and the Lawley–Hotelling trace are the

same, but their behavior under various violations of the null hypothesis and with small samples is different. Roy's largest root is different from the other three, even asymptotically.

None of the four multivariate criteria appears to be most powerful against all alternative hypotheses. For instance, Roy's largest root is most powerful when the null hypothesis of equal mean vectors is violated in such a way that the mean vectors tend to lie in one line within p-dimensional space. For most other situations, Roy's largest root performs worse than the other three statistics. Pillai's trace tends to be more robust to nonnormality and heteroskedasticity than the other three statistics.

The error sum of squares and cross products (SSCP) matrix is

$$E = A(Y'Y - B'X'XB)A'$$

and the SSCP matrix for the hypothesis is

$$H = A(CB)'\left\{C(X'X)C'\right\}(CB)A'$$

Let $\lambda_1 > \lambda_2 > \lambda_3 >$ s represent the nonzero eigenvalues of $E^{-1}H$.

$$s = min(p; vh)$$

, where p is the number of columns ofYA' (that is, the number of y variables or number of resultant transformed left-hand-side variables), and h is the hypothesis degrees of freedom.

Wilks's (1932) lambda statistic is

$$\Lambda = \prod_{i}^{s} \frac{1}{1+\lambda_i} = \frac{|E|}{|E+H|}$$

$$F = \frac{\left(1 - \Lambda^{\frac{1}{t}}\right) df_2}{\Lambda^{\frac{1}{t}} df_1}$$

and is a likelihood-ratio test. This statistic is distributed as the Wilks's distribution if E has the

Wishart distribution, H has the Wishart distribution under the null hypothesis, and E and H are

independent. The null hypothesis is rejected for small values of .

Pillai's (1955) trace is

$$V = \prod_{i}^{s} \frac{\lambda_i}{1+\lambda_i} = trace\left\{(E+H)^{-1}E\right\}$$

$$F = \frac{(2n+s+1)V}{(2m+s+1)(s-V)}$$

and the Lawley–Hotelling trace (Lawley 1938; Hotelling 1951) is

$$U = \sum_{i}^{s} \lambda_i = trace\left\{(E)^{-1}H\right\}$$

$$F = \frac{2(sn+1)U}{s^2(2m+s+1)}$$

and is also known as Hotelling's generalized T^2 statistic.

12.8 Canonical correlations

- Theory developed by Hotelling (1936, 1936)
- Correlating several metric dependent variables and metric independent variables simultaneously;
- Indentifying a subset of variables with the largest correlation from the set of many variables;
- Finding a linear combination among those variables that maximises correlation between them
- Do measurement of skulls and intelligence score elate to each other?
- Investment and profit?

Application in VAR model - cointegration analysis
Business example

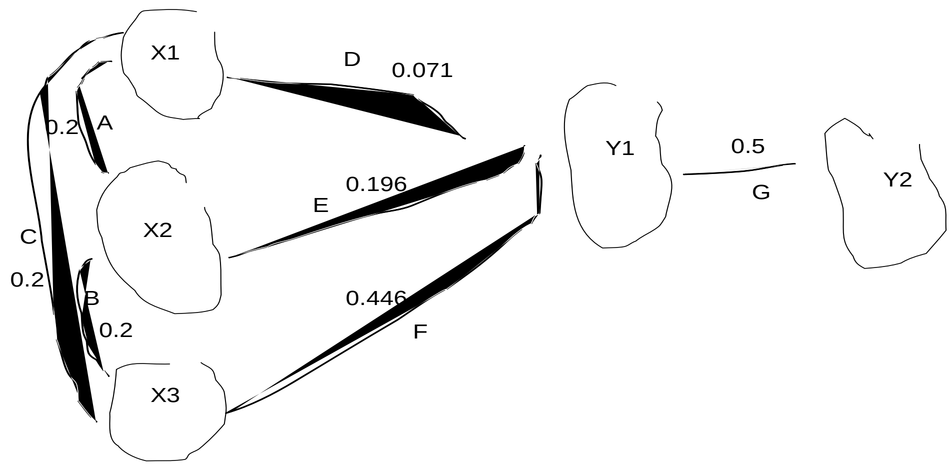
- Survey with 50 questions for a world class company and a particular company,
- Do correlations exists between the particular company and the world class company.

12.9 Structural equation modelling

- Explaining relationships among multiple variables; finding interrelationships
- Foundation on factor analysis and multiple regression analysis
- Constructs (exogenous variables) and latent factors
- Analysis of the covariance structure
- Model to define the entire set of relations
- Path diagrams
- Avoid spurious relations
- Simultaneous maximum likelihood estimation
  Example

$$\begin{array}{cc} & \begin{array}{ccc} x_1 & x_2 & y_1 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ y_1 \end{array} & \begin{bmatrix} 1 & & \\ 0.5 & 1 & \\ 0.6 & 0.7 & 1 \end{bmatrix} \end{array}$$

Price, service and Atmosphere are exogenous variables that lead to customer satisfaction. Customer satisfaction leads to the customer commitment.



Direct and indirect relation presente

$$\operatorname{cov}(X_1, X_2) = A = 0.2$$

$$\operatorname{cov}(X_2, X_3) = B = 0.2$$

$$\operatorname{cov}(X_1, X_3) = C = 0.2$$

$$\operatorname{cov}(X_1, Y_1) = D + AE + AF = 0.2$$

$$\operatorname{cov}(X_2, Y_1) = E + AD + BF = 0.3$$

$$\operatorname{cov}(X_3, Y_1) = F + BE + CD = 0.5$$

Equations for SEM Model

$$\operatorname{cov}(X_1, Y_1) = D + 0.2E + 0.2F = 0.2$$

$$\operatorname{cov}(X_2, Y_1) = 0.2D + E + 0.2F = 0.3$$

$$\operatorname{cov}(X_3, Y_1) = 0.2D + 0.2E + F = 0.5$$

Solving for Endogenous Path

$$\begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}\begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$$

$$\begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 1 \end{bmatrix}^{-1}\begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix}$$

$$\begin{bmatrix} D \\ E \\ F \end{bmatrix} = \begin{bmatrix} 1.071429 & -0.17857 & -0.17857 \\ -0.17857 & 1.071429 & -0.17857 \\ -0.17857 & -0.17857 & 1.071429 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.3 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.071428571 \\ 0.196428571 \\ 0.446428571 \end{bmatrix}$$

Interpretation of SEM Model

Customer satisfaction

$$Y_1 = 0.0714(\Pr ice) + 0.1964(Service) + 0.4464(Atmosphere)$$

Customer commitment

$$Y_2 = 0.5[0.0714(\Pr ice) + 0.1964(Service) + 0.4464(Atmosphere)] = 0.5[Y_1]$$

12.10 Steps in SEM modelling

- Defining individual construct
- Developing measurement models
- Empirical results
- Measurement model validity
- Specifying the structural models
- Assess SEM validity
- X1 attitude, X2 = environment x3 = job satisfaction ;0.5 = corr(x1x2) =A; corr(x1y1) =0.6= A+AC = 0.5+0.5C; corr(x1y1) =0.7= C+AB= C+0.5B;
  Solution: B = 0.33; C = 0.53

12.11 Factor analysis

- Condensing the underlying structure of the data
- Finding interrelation among the large number of variables in terms of a few integrated or latent factors.

- Total Variation
- By Individual specific factors
- By Errors

Variation due to common factors

$$(X_1, X_2, ..., X_n) = af + U + \epsilon$$

- Correlation matrix
- R-factor analysis/Q-factor analysis
- Sample size 10:1 to factors

Assumptions

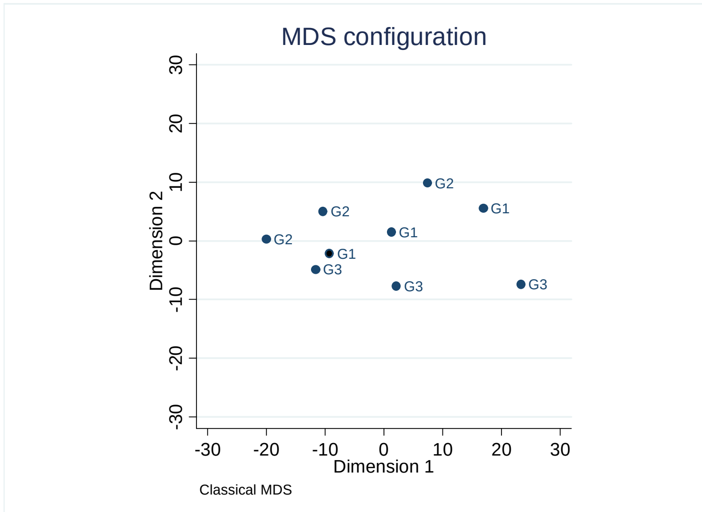
- Underlying structure exists; Multicollinearity among variables
- Bartlett test of sphericity; Measure of sampling adequacy
- Factor extraction : latent root criterion, scree test criteria, percentage of variation criteria
- Factor matrix : correlation; factor loading
- Factor rotation: orthogonal - quatrimax or varimax
- oblique : equimax
- Reliability: Cronbach's alpha;
- cross loading

12.12 Cluster analysis

- Grouping similar things together and finding the latent structure from more complex structure of multiple variables
- Initial cluster solution
- Minimum Euclidean distance and hierarchical procedure for cluster formation
- Measurement of distance:
- Euclidean distance

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

- Squared Euclidian distance
- City block
- Chebychev distance



- Mahalanobis distance

- Proximity mactrix and similarity index
- Graphical approach : Dandograms, single linkage, complete linkage, centroid method
- Non-heirarchical method: cluster seeds, sequential threshold, parallel threshold, optimisation

- Combination of both hierarchical and non-heirarchical methods.
- Stopping rule - agglomeration coefficient.

12.13 Multidimensional scaling

Indentifying the dimension underlying the respondents evaluations

Price, quality;

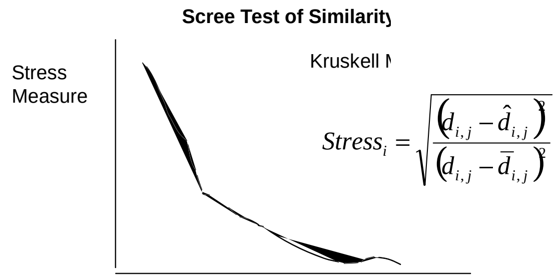
Perceptual map and similarity scale; Stress measures

12.14 Correspondence analysis (CA)

- Expected vs expected values

- Difference between and within rows and columns

- Similarity analysis - measure of similarity



$$d_{i,j} \qquad \hat{d}_{i,j} \qquad \bar{d}_{i,j}$$

- Optimal scaling or scoring
- Reciprocal averaging
- Homegeneity analysis

12.15 Analysis of Variance

One way ANOVA

Machines			
I	II	II	Mean
25.4	23.4	20	22.93333
26.31	21.8	22.2	23.43667
24.1	23.5	19.75	22.45
23.74	22.75	20.6	22.36333
25.1	21.6	20.4	22.36667

Total	124.65	113.05	102.95
Column mean	24.93	22.61	20.59

$$\bar{\bar{X}} = \frac{\sum_{j=1}^{c} \sum_{i=1}^{r} X_{ij}}{n}$$

22.71	Grand mean		

Correspondence analysis (CA): N

Actual Sales

	A	B	C	Total
Young	20	20	20	60
Adult	40	10	40	90
Old	20	10	40	70
Total	80	40	100	220

Expected Sales

	A	B	C	Total
Young	21.81818	10.90909	27.27273	60
Adult	32.72727	16.36364	40.90909	90
Old	25.45455	12.72727	31.81818	70
Total	80	40	100	220

Differences between actual and expe

	A	B	C	Total
Young	-1.818182	9.090909	-7.272727	0
Adult	7.272727	-6.363636	-0.909091	3.55E-15
Old	-5.454545	-2.727273	8.181818	0
Total	0	0	0	3.55E-15

Square of differences

	A	B	C	Total
Young	3.305785	82.64463	52.89256	138.843
Adult	52.89256	40.49587	0.826446	94.21488
Old	29.75207	7.438017	66.94215	104.1322
Total	85.95041	130.5785	120.6612	337.1901

Chi-Square (squared differences/expected freq)

	A	B	C	Total
Young	0.151515	7.575758	1.939394	9.666667
Adult	1.616162	2.474747	0.020202	4.111111
Old	1.168831	0.584416	2.103896	3.857143
Total	2.936508	10.63492	4.063492	17.63492

Chi-Square Test for corresponde

$$\chi^2_{i,n-1} = \frac{diffirence^2}{Expected\ value}$$

$$SST = \sum_{j=1}^{c}\sum_{i=1}^{r}\left(X_{ij} - \bar{\bar{X}}\right)^2$$

7.2361	0.4761	7.3441
12.96	0.8281	0.2601
1.9321	0.6241	8.7616
1.0609	0.0016	4.4521
5.7121	1.2321	5.3361
	58.2172	Total variation
47.164	Between Variation	
MSB	23.582	
df	c-1	

$$SSB = \sum_{i=1}^{r} n_j \left(X_{ij} - \bar{\bar{X}}\right)^2$$

0.2209	0.6241	0.3481
1.9044	0.6561	2.5921
0.6889	0.7921	0.7056
1.4161	0.0196	1E-04
0.0289	1.0201	0.0361
	11.0532	Within variation
MSW	0.9211	

$$SSW = \sum_{j=1}^{c}\sum_{i=1}^{r}\left(X_{ij} - \bar{X}_j\right)^2$$

F-Test	(MSB/MSW)	
25.602		

Two Way ANOVA

SST=SSFA+SSFB+SSAB+SSE

$$SST=\sum_{r=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}\left(X_{ijk}-\bar{\bar{X}}\right)^{2}=\sum_{r=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}X_{ijk}^{2}-\frac{\left(GT^{2}\right)}{rcn}$$

$$SSFA=cn\sum_{r=1}^{r}\left(\bar{X}_{i..}-\bar{\bar{X}}\right)^{2}=\sum_{r=1}^{r}\frac{X_{i..}^{2}}{cn}-\frac{\left(GT^{2}\right)}{rcn}$$

$$SSFB=rn\sum_{c=1}^{c}\left(\bar{X}_{.j.}-\bar{\bar{X}}\right)^{2}=\sum_{c=1}^{c}\frac{X_{..j.}^{2}}{rn}-\frac{\left(GT^{2}\right)}{rcn}$$

$$SSAB=n\sum_{r=1}^{r}\sum_{c=1}^{c}\left(X_{ij.}-\bar{X}_{i..}-\bar{X}_{.j.}-\bar{\bar{X}}\right)^{2}=\sum_{c=1}^{c}\frac{X_{..j.}^{2}}{rn}-\frac{\left(GT^{2}\right)}{rcn}=\sum_{r=1}^{r}\sum_{c=1}^{c}\frac{X_{i.j.}^{2}}{n}-\sum_{c=1}^{c}\frac{X_{..i.}^{2}}{cn}-\sum_{c=1}^{c}\frac{X_{..j.}^{2}}{rn}+\frac{\left(GT^{2}\right)}{rcn}$$

$$SSE=\sum_{r=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}\left(X_{ijk}-\bar{X}_{i,j}\right)^{2}=\sum_{r=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}X_{ijk}^{2}-\sum_{r=1}^{r}\sum_{c=1}^{c}\frac{X_{i.j.}^{2}}{n}$$

$$MSFB=\frac{SSAB}{(r-1)\,(c-1)}$$

$$MSFB=\frac{SSFB}{c-1}$$

$$MSFA=\frac{SSFA}{r-1}$$

$$MSFB=\frac{SSE}{rc\,(n-1)}$$

12.16 Multivariate Analysis of Variance (MANOVA)

MANOVA is similar to ANOVA and analysis of covariance (ANCOVA) except that arrays of characteristics are compared against various groups. Total variance is split between, within, interaction and error variances.

$$SS_{Total}=SS_{bg}+SS_{WG}$$

$$SS_{bg}=SS_{D}+SS_{T}+SS_{DT}$$

Consider a study on impacts of disability on scores of students within treatment and control groups.

$$SS_{Total}=SS_{D}+SS_{T}+SS_{DT}+SS_{S(DT)}$$

$$SST = \sum_i^r \sum_k^c \sum_m^n \left(X_{ijk} - \bar{\bar{X}}\right)^2 = n_k \sum_k \left(D_k - \bar{\bar{X}}\right)^2 + n_m \sum_k \left(T_m - \bar{\bar{X}}\right)^2$$

$$\left[n_{km} \sum_k \sum_m \left(D_k - \bar{\bar{X}}\right)^2 - n_k \sum_k \left(D_k - \bar{\bar{X}}\right)^2 - n_m \sum_k \left(T_m - \bar{\bar{X}}\right)^2\right] + \sum_i^r \sum_k^c \sum_m^n (X_{ijk} - DT_{km})^2$$

First part in the right hand side is total variance due to disability, second element total variance due to treatment, third element is for the interaction
In matrix notation

$$SST = \sum_i^r \sum_k^c \sum_m^n \left(X_{ijk} - \bar{\bar{X}}\right)\left(X_{ijk} - \bar{\bar{X}}\right)' = n_k \sum_k \left(D_k - \bar{\bar{X}}\right)\left(D_k - \bar{\bar{X}}\right)' + n_m \sum_k \left(T_m - \bar{\bar{X}}\right)\left(T_m - \bar{\bar{X}}\right)'$$

$$\left[n_{km} \sum_k \sum_m \left(DT_{km} - \bar{\bar{X}}\right)\left(DT_{km} - \bar{\bar{X}}\right)' - n_k \sum_k \left(D_k - \bar{\bar{X}}\right)\left(D_k - \bar{\bar{X}}\right)' - n_m \sum_k \left(T_m - \bar{\bar{X}}\right)\left(T_m - \bar{\bar{X}}\right)'\right]$$

$$+ \sum_i^r \sum_k^c \sum_m^n (X_{ijk} - DT_{km})(X_{ijk} - DT_{km})'$$

Example
Consider impacts of disability on scores in Math and English and IQ of students. Some of them are given special treatment (extra coaching) while others were not. Does treatment make difference in performance? How does the variance differ by groups and severity of disability?

		Mild			Moderate			Severe	
	MATH	ENGLISH	IQ	MATH	ENGLISH	IQ	MATH	ENGLISH	IQ
	115	108	110	100	105	115	89	78	99
Treatment	98	105	102	105	95	98	100	85	102
	107	98	100	95	98	100	90	95	100
	90	92	108	70	80	100	65	62	101
Control	85	95	115	85	68	99	80	70	95
	80	81	95	78	82	105	72	73	102

$$Y_{i11} = \begin{bmatrix} 115 \\ 108 \end{bmatrix} \begin{bmatrix} 98 \\ 105 \end{bmatrix} \begin{bmatrix} 107 \\ 98 \end{bmatrix}$$

- Test Wilk's lamda is ratio determinants:

$$\Lambda = \frac{|SS_{error}|}{|SS_{effects} + SS_{error}|}$$

See MANOVA.xls
Minitab Results from ANOVA: MS, ES, IQ versus DASB, Treat
Factor Type Levels Values
DASB fixed 3 Mild, Mod, Severe

- Treat fixed 2 T1, T2

Analysis of Variance for MS
Source DF SS MS F P
DASB 2 520.78 260.39 6.68 0.009
Treat 1 2090.89 2090.89 53.60 0.000
Error 14 546.11 39.01
Total 17 3157.78
S = 6.24563 R-Sq = 82.71% R-Sq(adj) = 79.00%

- Analysis of Variance for ES

Source DF SS MS F P
DASB 2 1126.78 563.39 13.32 0.001
Treat 1 1494.22 1494.22 35.33 0.000
Error 14 592.11 42.29
Total 17 3213.11
S = 6.50336 R-Sq = 81.57% R-Sq(adj) = 77.62%

- Analysis of Variance for IQ

Source DF SS MS F P
DASB 2 80.78 40.39 1.11 0.356
Treat 1 2.00 2.00 0.06 0.818
Error 14 507.67 36.26
Total 17 590.44
S = 6.02179 R-Sq = 14.02% R-Sq(adj) = 0.00%

- Means

Treat placeN MS ES IQ
T1 9 99.889 96.333 102.89
T2 9 78.333 78.111 102.22
DASB placeN MS ES IQ
Mild 6 95.833 96.500 105.00
Mod 6 88.833 88.000 102.83
Severe 6 82.667 77.167 99.83
MANOVA for DASB
s = 2 m = 0.0 n = 5.0
Test DF
Criterion Statistic F Num Denom P
Wilks' 0.22870 4.364 6 24 0.004
Lawley-Hotelling 3.34721 6.137 6 22 0.001
Pillai's 0.77711 2.754 6 26 0.033
placeCityRoy's 3.33961
SSCP Matrix for DASB
MS ES IQ
MS 520.8 761.7 203.39
ES 761.7 1126.8 301.61
IQ 203.4 301.6 80.78

SSCP Matrix for Error
MS ES IQ
MS 546.11 36.28 88.83
ES 36.28 592.11 354.50
IQ 88.83 354.50 507.67

- Partial Correlations for the Error SSCP Matrix

MS ES IQ
MS 1.00000 0.06380 0.16871
ES 0.06380 1.00000 0.64658
IQ 0.16871 0.64658 1.00000
EIGEN Analysis for DASB
Eigenvalue 3.3396 0.00760 0.00000
Proportion 0.9977 0.00227 0.00000
Cumulative 0.9977 1.00000 1.00000
Eigenvector 1 2 3
MS -0.02405 0.03595 -0.00458
ES -0.04512 -0.02405 0.01727
IQ 0.02605 -0.00128 -0.05294
MANOVA for Treat
s = 1 m = 0.5 n = 5.0
Test DF

Criterion Statistic F Num Denom P
Wilks' 0.10489 34.135 3 12 0.000
Lawley-Hotelling 8.53365 34.135 3 12 0.000
Pillai's 0.89511 34.135 3 12 0.000
placeCityRoy's 8.53365
SSCP Matrix for Treat
MS ES IQ
MS 2090.89 1767.56 64.667
ES 1767.56 1494.22 54.667
IQ 64.67 54.67 2.000
EIGEN Analysis for Treat
Eigenvalue 8.534 0.00000 0.00000
Proportion 1.000 0.00000 0.00000
Cumulative 1.000 1.00000 1.00000
Eigenvector 1 2 3
MS -0.03127 0.02873 -0.00938
ES -0.03976 -0.03414 0.01290
IQ 0.03228 0.00410 -0.04923

12.17 Principal components and factor analysis

- PCA analyses variance but the FA analyses covariance.
- Goal of PCA is to extract the maximum variance from the data set with a few orthogonal components
- FA aims to reproduce the correlation matrix with a few orthogonal factors.
- Take a correlation matrix R among numbers of variables.
- Find the Eigen values as
- $L = V'RV$ L is Eigen values, V is eigen vector and V' its transpose.
- Check $R = VLV'$ $VV' = I$

$$R = V\sqrt{L}\sqrt{L}V'$$

- Factor Loading matrix $A = V\sqrt{L}$ Orthogonal Rotation :

$$\Lambda = \begin{bmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{bmatrix}$$

$$A_{rotated} = A_{unrotated}\Lambda$$

see: principalcomponent.xls

Data

	cost	Lift	depth	powder
S1	32	64	65	67
S2	61	37	62	65
S3	59	40	45	43
S4	36	62	34	35
S5	62	46	43	40

R-Matrix

2.01636	1.9415	0.0044	

Eigen vectors		V	
0.352574	0.614235	0.662723	0.243321
-0.25131	-0.66369	0.676081	0.198157
-0.62731	0.322403	0.274838	-0.65345
-0.64731	0.279788	-0.16786	0.68886

Can use matrix routine in Minitab to compute
Transpose of Eigen Vectors

V'

0.352574	-0.25131	-0.62731	-0.64731
0.614235	-0.66369	0.322403	0.279788
0.662723	0.676081	0.274838	-0.16786
0.243321	0.198157	-0.65345	0.68886

L = V'RV	
2.02	0.00
0.00	1.94

It works so far.
Check on Check R = VLV'

Check V'V = I				VL				R - Approximately			
1.00	0.00	0.00	0.00		0.710885	1.192538		0.98	-0.97	-0.06	-0.13
0.00	1.00	0.00	0.00		-0.50671	-1.28857		-0.97	0.98	-0.10	-0.03
0.00	0.00	1.00	0.00		-1.26486	0.625962		-0.06	-0.10	1.00	0.99
0.00	0.00	0.00	1.00		-1.30519	0.543225		-0.13	-0.03	0.99	1.00

First two eigen vectors			Root of L		Factor A = V*rootL		
0.352574	0.614235		0.495943	0	cost	0.174857	0.304626
-0.25131	-0.66369		0	0.495943	Lift	-0.12464	-0.32915
-0.62731	0.322403				depth	-0.31111	0.159894
-0.64731	0.279788				powder	-0.32103	0.138759

Orthogonal Rotation
Factor Scores B = inv (R) A

Lamda rotation (transformation) matrix			
19			
0.988705	-0.14988	cosx	-sinx
0.149877	0.549772	sinx	cosx

A_unrottated * transformation mtrix

	Factor 1	Factor 2
cost	0.218538	0.141268
Lift	-0.17256	-0.16228
depth	-0.28363	0.134534
powder	-0.29661	0.124401

Inv (R)					Factor Scores B = inv (R) A	
25.48465	22.68874	-31.6546	35.47896		0.086583	0.156836
22.68874	21.38611	-24.8315	28.3122		-0.06195	-0.16958
-31.6546	-24.8315	99.9166	-103.95		-0.15415	0.082618
35.47896	28.3122	-103.95	109.5671		-0.15938	0.071194

Factor Extraction

Zcost = a11F1+a12F2	
Z_lift = a21F1+a1=22F2	
Z_powder = a41F1+a42F2	

12.18 Difference between Principal component analysis (PCA) and factor analysis (FA)

PCA analyses variance but the FA analyses covariance. Goal of PCA is to extract the maximum variance from the data set with a few orthogonal components while FA aims to reproduce the correlation matrix with a few orthogonal factors.

12.19 Canonical correlations

Hotelling (1936, 1935): the maximum correlations between linear functions of the two vector variables. Consider z1 vector of variables with p1 elements and z2 with p2 elements. Then square correlation matrix R of p1+p2 is

Take a correlation matrix:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

$$R = \begin{vmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{vmatrix} = R_{11}R_{22} - R_{12}R_{21} = 0$$

$$R_{22} = R_{12}R_{11}^{-1}R_{21}$$

$$I = R_{22}^{-2}R_{21}R_{11}^{-1}R_{12}$$

For instance, let p1= 2 and p2=2;n then

$$R = \begin{bmatrix} 1 & 0.4 & 0.5 & 0.6 \\ 0.4 & 1 & 0.3 & 0.4 \\ 0.5 & 0.3 & 1 & 0.2 \\ 0.6 & 0.4 & 0.2 & 1 \end{bmatrix}$$

N=100.

$$R_{11} = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$$

;

$$R_{22} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$$

$$R_{12} = R_{21} = \begin{pmatrix} 05 & 0.3 \\ 0.6 & 04 \end{pmatrix}$$

$$I = R_{22}^{-2}R_{21}R_{11}^{-1}R_{12} = \begin{pmatrix} 1.041 & -0.208 \\ 0.208 & 1.041 \end{pmatrix} \begin{pmatrix} 05 & 0.3 \\ 0.6 & 04 \end{pmatrix} \begin{pmatrix} 1.190 & -0.476 \\ -0.476 & 1.190 \end{pmatrix} \begin{pmatrix} 05 & 0.6 \\ 0.3 & 04 \end{pmatrix}$$

		0.250992	
0.277777778	0.340278		

$$D = \begin{vmatrix} 0.206 - \lambda & 0.251 \\ 0.277 & 0.341 - \lambda \end{vmatrix} = 0;$$

$$\lambda^2 - 0.547\lambda + 0.003 = 0$$

$\lambda_1 = 0.546$; $\lambda_2 = 0.001$; Notice that sum of the roots equals trace of the matrix.

Two canonical correlations

$$R_{C1} = \sqrt{\lambda_1 = 0.546} = 0.74$$

and

$$R_{C2} = \sqrt{\lambda_2 = 0.001} = 0.03$$

placeCityBartlett's test and Wilks' Lamda

$$\Lambda = (1 - \lambda_1)(1 - \lambda_2) = (1 - 0.546)(1 - 0.001) = 0.454$$

$$\chi^2 = -\left[n - 0.5\left(p1 + p2 + 1\right)\right] \ln \lambda_1 = -\left[99 - 0.5\left(2 + 2 + 1\right)\right] \ln 0.454 = 77$$

df =4, reject that z1 and z2 are unrelated as

$$p\left(\chi^2 = 77\right) > 0.999$$

.

After removing the first root the next largest root is

$$\Lambda' = (1 - \lambda_2) = (1 - 0.001) = 0.999$$

$\chi^2 = -0.965$. Only the first root is significant.
Find the eigenvector corresponding to the first root

$$M - \lambda I = \begin{bmatrix} 0.206 - \lambda & 0.251 \\ 0.277 & 0.341 - \lambda \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

$$M - \lambda I = \begin{bmatrix} 0.206 - 0546 & 0.251 \\ 0.277 & 0.341 - 0.546 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} or \begin{bmatrix} -0.340 & 0.251 \\ 0.277 & -0.205 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

cofactor of the first row is the canonical coefficient

$$v = \begin{pmatrix} 0.205 & -0.278 \end{pmatrix}$$

.

The variance of this factor is

$$vR_{22}v' = \begin{pmatrix} -0.205 & -0.278 \end{pmatrix} \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \begin{pmatrix} -0.205 \\ -0.278 \end{pmatrix} = \theta$$

Corresponding Eigen vector

$$d = v\theta^{-\frac{1}{2}} = \begin{pmatrix} 0.545 \\ 0.737 \end{pmatrix}$$

Now

$$C = \frac{R_{11}^{-1}R_{12}d_j}{\sqrt{\lambda_i}} = \begin{pmatrix} 1.190 & -0.476 \\ -0.476 & 1.190 \end{pmatrix} \begin{pmatrix} 05 & 0.6 \\ 0.3 & 04 \end{pmatrix} \begin{pmatrix} 0.546 \\ 0.737 \end{pmatrix} \frac{1}{\sqrt{0.546}} = \begin{pmatrix} 0.856 \\ 0.278 \end{pmatrix}$$

Thus the extracted variables with unit variance and correlation of 0.74 are

$$0.856z_{11,i} + 0.278z_{12,i} = x_i$$

$$0.545z_{21,i} + 0.737z_{22,i} = y_i$$

What is the correlation of new canonical variates with the original variables?

$$s_1 = R_{11}c = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix} \begin{pmatrix} 0.856 \\ 0.278 \end{pmatrix} = \begin{pmatrix} 0.967 \\ 0.620 \end{pmatrix}$$

$$s_2 = R_{22}d = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \begin{pmatrix} 0.545 \\ 0.737 \end{pmatrix} = \begin{pmatrix} 0.692 \\ 0.846 \end{pmatrix}$$

Proportion of the left side battery extracted by the first canonical covariate

$$\frac{s_1' s_1}{p1} = \begin{pmatrix} 0.967 & 0.620 \end{pmatrix} \begin{pmatrix} 0.967 \\ 0.620 \end{pmatrix} \frac{1}{2} = 0.660$$

$$\frac{s_2' s_2}{p2} = \begin{pmatrix} 0.692 & 0.846 \end{pmatrix} \begin{pmatrix} 0.692 \\ 0.846 \end{pmatrix} \frac{1}{2} = (1.19458)\frac{1}{2} = 0.597$$

References

[1] Cooley WW and Lohnes PR (1971) Multivariate Data Analysis, John Wiley and Sons.

Reading a matrix and getting Eigen values in Minitab
See canonical.xls
Take a data

7	4	3
4	1	8
6	3	5
8	6	1
8	5	7
7	2	9
5	3	3
9	5	8
7	4	5
8	2	2

Correlation matrix is
1.000.67-0.10

0.67	1.00	-0.29
-0.10	-0.29	1.00

12.20 Discriminant Analysis

Statistical technique to compare the means of a set of independent variables for two or more groups. Males differ from female, North- South, skilled vs unskilled.

- Categorical dependent variable
- Logistic regression
- Frequency distributions for two/multiple of groups.
- Steps for modelling discriminant analysis
- Sort observations by groups and check the differences in group means
- Determine independent variables and sample size
- Assumptions - normality, linearity, no multicollinearity, equal dispersions
- Estimation of the characteristic function - simultaneous or stepwise

- Check the significance of the discriminant function
- Assess prediction accuracy
- Interpret results - discriminant weights, loadings, partial fractions, split samples and cross validation
  Examples

Fundamental equation:

$$S_{total} = S_{bg} + S_{wg}$$

Similar to manova.

Linear discriminant function:

$$D_i = d_{i1}z_1 + d_{i2}z_2 + ... + d_{ip}z_p$$

Classification function:

$$C_j = C_{j0} + c_{j1}X_1 + c_{j2}X_2 + ... + c_{jp}X_p$$

$C_j = W^{-1}M_j$;

$$c_{j0} = \left(-\frac{1}{2}\right) C_j' M_j$$

See discriminant.xls.

Example

IQ	Info	Verbal	Age	Group
87	5	31	6.4	G1
97	7	36	8.3	G1
112	9	42	7.2	G1
102	16	45	7	G2
85	10	38	7.6	G2
76	9	32	6.2	G2
120	12	30	8.4	G3
85	8	28	6.3	G3
99	9	27	8.2	G3

- Between group variation

SSCP Matrix (adjusted) for Group
IQ Info Verbal Age
IQ 314.9 -71.56 -180.0 14.49
Info -71.6 32.89 8.0 -2.22
Verbal -180.0 8.00 168.0 -10.40
Age 14.5 -2.22 -10.4 0.74

- Within group variation

SSCP Matrix (adjusted) for Error
IQ Info Verbal Age
IQ 1286.00 220.000 348.333 50.000
Info 220.00 45.333 73.667 6.367
Verbal 348.33 73.667 150.000 9.733
Age 50.00 6.367 9.733 5.493

- Tests

MANOVA for Group
s = 2 m = 0.5 n = 0.5

- Test DF
  Criterion Statistic F Num Denom P
  Wilks' 0.01048 6.577 8 6 0.017
  Lawley-Hotelling 19.07513 4.769 8 4 0.074
  Pillai's 1.77920 8.058 8 8 0.004
  Roy's 13.48590

- Pooled Group Variance-Covariance Matrix (W)

Pooled Covariance Matrix
IQ Info Verbal Age
IQ 214.333
Info 36.667 7.556
Verbal 58.056 12.278 25.000
Age 8.333 1.061 1.622 0.916

Inverse of W
Inverse of the pooled matrix

	IQ	Info	Verbal	Age
IQ	0.043595	-0.20182	0.009539	-0.17972
Info	-0.20182	1.628831	-0.37056	0.605476
Verbal	0.009539	-0.37056	0.200678	-0.0129
Age	-0.17972	0.605476	-0.0129	2.048136

Group Means

	Pooled means for group			
Variable	Mean	G1	G2	G3
IQ	95.889	98.667	87.667	101.333
Info	9.4444	7	11.6667	9.6667
Verbal	34.333	36.333	38.333	28.333
Age	7.2889	7.3	6.9333	7.6333

Classification coefficients

	G1: Mem	G2: perc	G3: Comm
IQ	1.923305	0.586914	1.365133
Info	-17.5547	-8.69661	-10.5828
Verbal	5.544256	4.116101	2.971779
Age	0.988946	5.014559	2.910278

Standardised score equals

Classification scores

C1	119.9839
C2	95.99969
C3	105.3801

This student will be assigned to group 1 because of highest score in it.

12.21 Cluster Analysis

- Grouping similar things together and finding the latent structure from more complex structure of multiple variables
- Initial cluster solution
- Minimum Euclidean distance and hierarchical procedure for cluster formation
- Measurement of distance:
- Euclidean distance or Squared Euclidian distance
- City block
- Chebychev distance
- Mahalanobis distance
- Proximity mactrix and similarity index

- Graphical approach : Dandograms, single linkage, complete linkage, centroid method
- Non-heirarchical method: cluster seeds, sequential threshold, parallel threshold, optimisation
- Combination of both hierarchical and non-heirarchical methods.
- Stopping rule - agglomeration coefficient.
- 1.Correlation
- 2.similarity index based on Euclidian Distance Measure
- 3.Partition the variables by proximity see the proximity matrix
- 4.Dendogram, Centroid methods
- Similarity measure

$$-1 \leq s_{i,j} \leq 1$$

- Clustering criteria T = W + B

T = total distance, W = within group distance, B = between group distance.

$$T = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n} (x_{i,j} - \bar{x})(x_{i,j} - \bar{x})'$$

$$W = \frac{1}{n-g} \sum_{i=1}^{g} \sum_{j=1}^{n} (x_{i,j} - \bar{x}_j)(x_{i,j} - \bar{x}_i)'$$

$$T = \frac{1}{n} \sum_{i=1}^{g} (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})$$

- Choose the cluster that minimises the sum of within distance and raises between distance.

Minimising the trace of Euclidian distance:

$$E = \sum_{i=1}^{n} d_{i,c(i)}^2$$

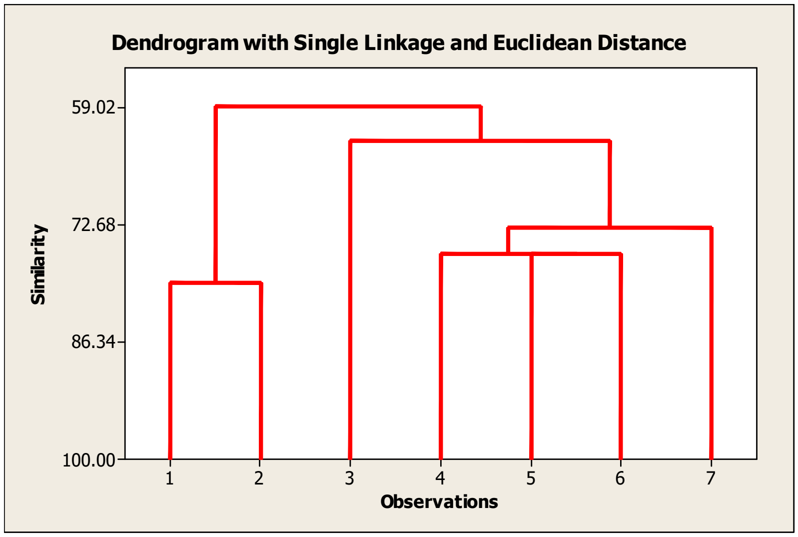
Minimisation of determinant of W

Maximisation of the trace of BW^{-1} or

$$\frac{\det(T)}{\det(W)}$$

which can be expressed in terms of eigenvalues, λ_i: Here

$$trace\left(BW^{-1}\right) = \sum_{i=1}^{p} \lambda_i$$



and

$$\frac{\det(T)}{\det(W)} = \prod_{i=1}^{p} (1 + \lambda_i)$$

References

[1] Anderson T.W. (1996) R.A. Fisher and Multivariate Analysis, Statistical Science, 11:1:20-34.

[2] Bartlett M.S. (1947) Multivariate Analysis, Supplement to the Journal of the Royal Statistical Society, 9:2:176-197.

[3] Beyer A, REA Farmer, J Henry and M. Marcellino (2005) Factor Analysis in a New Keynesian Model, European Central Bank, Working Paper Series, no. 510, August

[4] Chung KC, Tan SS. Holdsworth DK (2008) Insolvency Predication Model Using Multivariate Discriminat Analysis and Artificial Neuroal Network for the Finance Industry in New Zealand, International Journal of Business Management, 3:1:19-28.

[5] Finnerty JE (1976) Insiders' Activity and Inside Information: A Multivariate Analysis, the Journal of Financial and Quantitative Analysis, 11:2:205-215.

[6] Gatty R (1966) Multivariate analysis for marketing research: An Evaluation, Applied Statistics, 15:3:157-172.

[7] Gudmundsson G. (1977) Multivariate analysis of economic variables, Applied Statistics, 26:1:48-59.

[8] Hirschberg JG, placeE Maasoumi and DJ Slottje (2001) Clusters of Attributes and Well-Being in the placecountry-regionUSA, Journal of Applied Econometrics, 16:3:445-460.

[9] King LJ (1970) Discriminant Analysis: A Review of Recent Theoretical Contributions and Applications, Economic Geography, 46: Supplement:367-378.

[10] Luzzi GF, YFluckiger and S Weber (2006) A cluster analysis of multidimensional poverty in placecountry-regionSwitzerland, Haute Ecole de Gestion de Geneve.

[11] Maasoumi E (1986) The measurement and decomposition of multi-dimensional inequality, Econometrica, 54:4:991-997.

[12] Pinches GE and K A Mingo (1973) A Multivariate Analysis of Industrail Bond Ratings, Journal of Financce, 28:1:1-18.

[13] Proudfoot MJ (1937) City Retail Structure, Economic Geography, 13:4:425-428.

[14] Rao C R (1972) Recent Trends of Research Work in Multivariate Analysis, Biometrics- Special Multivariate Issue, 28:1:3-22.

[15] Rayner A C (1970) The Use of Multivariate Analysis in Development Theory: A Critique of the Approach Adopted by Adelman and Morris, Quarterly Journal of Economics, 84:4:639-647.

[16] Rider P R (1936) Annual Survey of Statistical Technique: Developments in the Analysis of Multivariate Data - Part I, Econometrica, 4:3:264-268.

[17] Schervish MJ (1987) A Review of Multivariate Analysis, Statistical Science, 2:4:396-413.

[18] Sorensen CK and JMP Gutierrez (2006) Euro Area Banking Sector Integration Using Hierarchical Cluster Analysis Techniques, European Central Bank, Working Paper Series, no. 626, May.

[19] Stevens DL (1973) Financial Characteristics of Merged Firms: A Multivariate Analysis, Journal of Financial and Quantitative Analysis, 8:2:149-158.

[20] Thomas BAM (1961) Some Industrial Applications of Multivariate Analysis, Applied Statistics, 10:1:1-8.
Wishart J (1955) Multivariate Analysis, Applied Statistics, 4:2:103-116.

[21] Cooley WW and Lohnes PR (1971) Multivariate Data Analysis, John Wiley and Sons.

[22] Hair JF, WC Black, BJ Babin, RE Anderson and RL Tatham (2006) Multivariate Data Analysis, Sixth Edition, Pearson International.

[23] Tabachnick BG and L S Fidell (2007) Using multivariate Statistics, Person.

13 L11: Forecasting

Business economists have to make decisions in circumstances where the future values of many economic variables are unknown. These include income and consumption, saving and investment, government revenue and spending, imports and exports, sales revenue and profit, prices and supplies of goods and services and factors of production, interest rate and exchange rates, rates of growth of output, employment and the capital stocks, rates of inflation and unemployment, values of stocks and bonds and many other financial and real assets. These uncertainties may partly arise due to changes in equilibrium in various markets due to changes in preferences of consumers, technology of producers and policies of the government or corporate sectors and strategies taken by them. Unpredictable natural factors such as climates, earthquakes or floods, outbreak of epidemics or the international factor such as wars, changes in the rules and regulations of trade and payment as well as in the mobility of factors of production may alter the probability distributions, the data generating processes, of these economic variables. Economic agents -households, firms, government and traders- nevertheless have to take decisions on the basis of forecasts or predication or simulations of those variables based upon available information though it is almost impossible to forecast their values exactly beforehand.

Economists use models to consolidate their reasoning and to justify their arguments or predict or forecast about future economic activities. These models that are taken scientific tools to analyse economic issues objectively can be small or big, static or dynamic, deterministic or stochastic, strategic or optimising or analytic, mathematical or econometric. They may analyse aggregate behaviour at macro level or more details in the micro level. They may aim to explain local, regional, national or global markets or economic policies aimed for stability, efficiency, growth and redistribution.

These models can be grouped into two major categories. First type of model mainly relies on mathematical structure for describing the economy. These relations are expressed generally in terms of algebra and involve solving simultaneous equations with various optimising conditions that are used to study maximising or minimising behaviour of consumer or producers in the economy. General equilibrium models or strategic models fit into this category. Solutions of these models are defined by a set of parameters. Then there are econometric models that aim to fit the model as accurately as possible to the observed data set of random variables, these models aim to minimise the errors of prediction. These models apply stochastic probability theories to theoretical economic models and aim to study empirical facts relating to an economic issue. Both of these economic models can serve as laboratory for analysis of various instruments of economic policy before they are applied to real world situation.

Literature in forecasting is evolving very rapidly:

- Bradford and Kelejian (1977), Min and Zellner (1993), Hamilton and Lin (1996), Hendry (1997), Ericsson and Marquez (1998), Clements and Krolzig(1998),

- Harvey, Leybourne and Newbold (2001), Artis and Marcellino (2001),Stock and Watson (2002) Ng and Vogelsang (2002), Clements and Hendry(2002), Gabriel and Martins (2004),

- Hendry, and Clements (2004), Koop and Potter(2004, 2007),

- Elliott, Timmermann and and Komunjer (2005), Pesaran , Pettenuzzo, and Timmermann (2006), Koop and Simon Potter (2007) Bai and Ng (2008),Rossi and Sekhposyan (2011), Engle and Sokalska (2012a,b), Fawcett, Kapetanios, Mitchell and Price (2013), Smith and Velázquez (2013)

- Texts: Granger and Newbold (1986), Judge, Hill, Griffiths, Lutkepohl and Lee (1988) Harvey (1989) Harvey (1990) Griffiths , Hill and Judge (1993), Hamilton (1994), Hendry (1995) Campbell , Lo and MacKindlay (1997),Davidson (2000), Holly and Weale ed. (2000), Clement and Hendry ed. (2002), Harris and Sollis (2003), Mills (2000, 2003), Davidson and MacKinnon (2004), Blundell, Newey and Persson (2006), Singleton (2006), Koop (2008), Enders (2010)

13.1 What is forecasting?

- What is a forecast?
- It is a statement about the future.
- How are forecasts made?
- There are many ways: guess, coin tossing, qualitative judgement, sophisticated economic models.
- What is the best forecast? Which is close to the truth.
- Why do most forecasts fail?
- Future is uncertain.

Econometric Models for Forecasting

- Single equation models
- Cross section predictions (behavioural analysis; surveys)
- Linear and nonlinear time series models: AR, ARMA, ARCH-GARCH, CHARMA. TAR, STAR (smooth transition AR); forecasting volatility, artificial neural network, statespace, Kalman Filter and markov switching models
- Simultaneous equation models/structural equation models
- Prediction and forecasts; forecast intervals; survey forecasts
- Rational expectations; leading indicators; Delphi method
- Multivariate timeseries VAR Models
- Ramsey models: Preferences, technology, prices, process of economy over years
- Neoclassical growth model
- Stochastic general equilibrium models
- New classical innovations on technology shocks
- Applied dynamic general equilibrium models
- Large scale models for regional, national and global economy
- Strategic Models –predictions from games - Nash Bargaining, repeated games, signalling, screening, cooperative and non-cooperative games of political economy; corporate strategies, trade negotiations

13.1.1 Principles of Forecasting:

Minimum Sum Square (MSE, MAE, RMSE)

$$MSE = E\left(Y_{t+1} - Y^*_{t+1/t}\right)^2 \tag{1626}$$

Smallest MSE is obtained in prediction of Y_{t+1} is conditional on all of its past values.

$$Y^*_{t+1/t} = E\left(Y_{t+1}/X_t\right) \tag{1627}$$

Here $X_t : Y_t, Y_{t-1}, ..., Y_{t-m+1}$
Linear projection

$$Y^*_{t+1/t} = \alpha' X_t \tag{1628}$$

$$E\left[\left(Y_{t+1} - \alpha' X_t\right) X_t'\right] = 0 \tag{1629}$$

$$\alpha' = E\left[Y_{t+1} X_t'\right] \left[E\left(X_t X_t'\right)\right]^{-1} \tag{1630}$$

13.1.2 Simple extrapolation:

Crude Morving Average Projections
Weekly updates

$$\overline{z}_W = \frac{1}{4}\left(z_1 + z_2 + z_3 + z_4\right) \tag{1631}$$

Quarterly

$$\overline{z}_Q = \frac{1}{4}\left(z_1 + z_2 + z_3 + z_4\right) \tag{1632}$$

Monthly

$$\overline{z}_M = \frac{1}{12}\left(z_1 + z_2 + z_3 + z_4 + ... + z_{12}\right) \tag{1633}$$

$$\widehat{\mu} = \sum_{i=-n}^{n} a_i y_{t-i}; \qquad i = 0, \pm 1, \pm 2, ..., \pm n \tag{1634}$$

13.1.3 Components of time sereis

Components of a series (trend, season, cycle and random)

$$y_t = \mu_t + \gamma_t + \psi_t + \nu_t + \epsilon_t \qquad \epsilon_t \sim NID\left(0, \sigma^2_{\epsilon_t}\right) \tag{1635}$$

Stochastic trend

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \qquad \eta_t \sim NID\left(0, \sigma^2_{\eta_t}\right) \tag{1636}$$

$$\beta_t = \beta_{t-1} + \zeta_t \qquad \zeta_t \ \sim \ NID\left(0, \sigma^2_{\zeta_t}\right) \tag{1637}$$

Seasonal dummies

$$\gamma_t = -\gamma_{t-1} - ... - \gamma_{t-s+1} + \omega_t \qquad \omega_t \ \sim \ NID\left(0, \sigma^2_{\omega_t}\right) \tag{1638}$$

Refer to STAMP mannual (p.140)
Seasonal factor in dummy or trigonometric form
or trigonometric seasonal representation

$$\gamma_t = \sum_{j=-n}^{s/2} \gamma_{j,t} \tag{1639}$$

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma^*_{j,t} \end{pmatrix} = \begin{pmatrix} \cos\lambda_j & \cos\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma^*_{j,t-1} \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega^*_{j,t} \end{pmatrix} \qquad \begin{matrix} j = 1, ..s/2 \\ t = 1, ...T \end{matrix} \tag{1640}$$

$\lambda_j = \frac{2\pi j}{s}$ is frequency in radians.
Cyclical and random components
$\lambda_j = 2\pi j`$.
Cycle ψ_t

$$\begin{pmatrix} \psi_{j,t} \\ \psi^*_{j,t} \end{pmatrix} = \rho_\psi \begin{pmatrix} \cos\lambda_j & \cos\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \psi_{j,t-1} \\ \psi^*_{j,t-1} \end{pmatrix} + \begin{pmatrix} \kappa_{j,t} \\ \kappa^*_{j,t} \end{pmatrix} \qquad \begin{matrix} j = 1, ..s/2 \\ t = 1, ...T \end{matrix} \tag{1641}$$

$0 < \rho_\psi \leq 1$ is called damping factor $\kappa_{j,t}$ and $\kappa^*_{j,t}$ are NID disturbance terms.$NID\left(0, \sigma^2_{\kappa_t}\right)$
Randome error evolves according to

$$\nu_t = \rho_\psi \nu_{t-1} + \xi_t ....\xi_t \sim NID\left(0, \sigma^2_{\xi_t}\right) \tag{1642}$$

13.2 Forecasting from Random Walk

Random Walk

$$y_t = y_{t-1} + \epsilon_t \tag{1643}$$

Starting from initial condition: $y_1 = y_0$ for $t = 1$.

$$y_1 = y_0 + \epsilon_1 \tag{1644}$$

$$y_2 = y_1 + \epsilon_2 = y_0 + \epsilon_1 + \epsilon_2 \tag{1645}$$

$$y_5 = y_4 + \epsilon_5 = y_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 + \epsilon_5 \tag{1646}$$

$$y_t = y_0 + \sum_{t=0}^{t} \epsilon_i \tag{1647}$$

Random Walk

$$E(y_t) = y_0 + \sum_{t=0}^{t} E(\epsilon_i) = y_0 \tag{1648}$$

$$var(y_t) = t\sigma^2 \tag{1649}$$

$$var(y_{t-s}) = (t-s)\sigma^2 \tag{1650}$$

Random Walk process is non-stationary.
Forecast function

$$E_t y_{t+1} = y_t \tag{1651}$$

Random walk with a drift

$$y_1 = y_0 + a_0 + \epsilon_1 \tag{1652}$$

Random Walk

$$y_2 = y_1 + a_0 + \epsilon_2 = y_0 + a_0 + \epsilon_1 + \epsilon_2 \tag{1653}$$

$$y_5 = y_4 + \epsilon_5 = y_0 + 4a_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 + \epsilon_5 \tag{1654}$$

$$y_t = y_0 + ta_0 + \sum_{t=0}^{t} \epsilon_i \tag{1655}$$

Forecast function

$$y_{t+s} = y_0 + (t+s)\,a_0 + \sum_{i=1}^{t+s} \epsilon_i \tag{1656}$$

Random Walk

$$y_{t+s} = y_t + sa_0 + \sum_{i=1}^{s} \epsilon_{t+i} \tag{1657}$$

Forecast now contains deterministic trend term sa_0 in addition to the pure random walk

$$E_t y_{t+s} = y_t + sa_0 \tag{1658}$$

13.2.1 AR(1) forecast

$$y_t = \delta + \theta_1 y_{t-1} + e_t \tag{1659}$$

h =1 ahead Forecast

$$y_{T+1} = \delta + \theta_1 y_T + e_{T+1} \quad e_{T+1} \sim N(0,1) \tag{1660}$$

Mean forecast:

$$\widehat{y}_{T+1} = E\left(y_{T+1}\right) = \delta + \theta_1 y_T \tag{1661}$$

Estimate of Forecast error

$$\widehat{e}_{T+1} = y_{T+1} - \widehat{y}_{T+1} = \delta + \theta_1 y_T + e_{T+1} - \delta - \theta_1 \widehat{y}_T \tag{1662}$$

Variance of h =1 Forecast error

$$var\left(\widehat{e}_{T+1}\right) = \sigma_e^2 \tag{1663}$$

h =2 ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{T+1} + e_{T+1} \quad e_{T+2} \sim N(0,1) \tag{1664}$$

Mean forecast:

$$\widehat{y}_{T+2} = E\left(y_{T+2}\right) = \delta + \theta_1 y_{T+1} \tag{1665}$$

Estimate of Forecast error

$$\begin{aligned} \widehat{e}_{T+2} &= y_{T+2} - \widehat{y}_{T+2} = \delta + \theta_1 y_{T+1} + e_{T+2} - \delta - \theta_1 \widehat{y}_{T+1} \\ &= e_{T+2} + \theta_1\left(y_{T+1} - \widehat{y}_{T+1}\right) = e_{T+2} + \theta_1 e_{T+1} \end{aligned} \tag{1666}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+2}\right) = \sigma_e^2\left(1 + \theta_1^2\right) \tag{1667}$$

h period ahead Forecast

$$y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h} \quad e_{T+h} \sim N(0,1) \tag{1668}$$

Mean forecast:

$$\widehat{y}_{T+h} = E\left(y_{T+h}\right) = \delta + \theta_1 \widehat{y}_{T+h-1} \tag{1669}$$

Estimate of Forecast error

$$\begin{aligned} \widehat{e}_{T+h} &= y_{T+h} - \widehat{y}_{T+2} = \delta + \theta_1 y_{T+h-1} + e_{T+h} - \delta - \theta_1 \widehat{y}_{T+h-1} \\ &= e_{T+h} + \theta_1\left(y_{T+h-1} - \widehat{y}_{T+h-1}\right) = e_{T+h} + \theta_1 e_{T+h-1} \end{aligned} \tag{1670}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+h}\right) = \sigma_e^2\left(1 + \theta_1^2 + \theta_1^2 + ... + \theta_1^{2(h-1)}\right) \tag{1671}$$

13.2.2 MA(1) forecast

Forecast with MA(1)

$$y_t = \mu + e_t + \alpha_1 e_{t-1} \tag{1672}$$

h=1 period ahead forecast

$$y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T \tag{1673}$$

Mean forecast

$$E\left(y_{T+1}\right) = \widehat{y}_{T+1} = \mu + \alpha_1 e_T \tag{1674}$$

Forecast error

$$y_{T+1} - \widehat{y}_{T+1} = \mu + e_{T+1} + \alpha_1 e_T - \mu - \alpha_1 e_{T+1} = e_{T+1} \tag{1675}$$

Variance of forecast:

$$var\left(y_{T+1} - \widehat{y}_{T+1}\right) = var\left(e_{T+1}\right) = \sigma_e^2 \tag{1676}$$

h=2 period ahead Forecast

$$y_{T+2} = \mu + e_{T+2} + \alpha_1 e_{T+1} \tag{1677}$$

Mean forecast

$$E\left(y_{T+2}\right) = \widehat{y}_{T+2} = \mu \tag{1678}$$

Forecast error

$$y_{T+2} - \widehat{y}_{T+2} = \mu + e_{T+2} + \alpha_1 e_{T+1} - \mu = e_{T+2} + \alpha_1 e_{T+1} \tag{1679}$$

$$var\left(y_{T+2} - \widehat{y}_{T+2}\right) = var\left(e_{T+2}\right) = var\left(e_{T+2} + \alpha_1 e_{T+1}\right) = \sigma_e^2\left(1 + \alpha_1^2\right) \tag{1680}$$

Similarly mean and variance of h period ahead forecast:

$$y_{T+h} = \mu + e_{T+h} + \alpha_1 e_{T+h-1} \tag{1681}$$

$$E\left(y_{T+h}\right) = \widehat{y}_{T+h} = \mu \tag{1682}$$

Forecast error

$$y_{T+h} - \widehat{y}_{T+h} = \mu + e_{T+h} + \alpha_1 e_{T+h-1} - \mu = e_{T+h} + \alpha_1 e_{T+h-1} \tag{1683}$$

$$var\left(y_{T+2} - \widehat{y}_{T+2}\right) = var\left(e_{T+h}\right) = var\left(e_{T+h} + \alpha_1 e_{T+h-1}\right) = \sigma_e^2\left(1 + \alpha_1^2\right) \tag{1684}$$

13.2.3 ARMA(1,1) forecast

Forecasts using ARMA(1,1) process:

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1} \tag{1685}$$

h=1 period ahead Forecast

$$y_{T+1} = \delta + \theta_1 y_{t-1} + e_{T+1} + \alpha_1 e_T \tag{1686}$$

Mean forecast

$$E\left(y_{T+1}\right) = \widehat{y}_{T+1} = \delta + \theta_1 y_{t-1} + \alpha_1 e_T \tag{1687}$$

Forecast error

$$\begin{aligned} \widehat{e}_{T+1} &= \left(y_{T+h} - \widehat{y}_{T+h}\right) = \\ \delta + \theta_1 y_{t-1} + e_{T+1} + e_{T+1} + \alpha_1 e_T - \delta - \theta_1 y_{t-1} - \alpha_1 e_T &= e_{T+1} \end{aligned} \tag{1688}$$

Forecast error

$$\begin{aligned} \widehat{e}_{T+1} &= \left(y_{T+h} - \widehat{y}_{T+h}\right) = \delta + \theta_1 y_{t-1} + e_{T+1} + e_{T+1} \\ +\alpha_1 e_T - \delta - \theta_1 y_{t-1} - \alpha_1 e_T &= e_{T+1} \end{aligned} \tag{1689}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+1}\right) = var\left(y_{T+h} - \widehat{y}_{T+h}\right) = var\left(e_{T+1}\right) = \sigma_e^2 \tag{1690}$$

$$y_t = \delta + \theta_1 y_{t-1} + e_t + \alpha_1 e_{t-1} \tag{1691}$$

h=2 period ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1} \tag{1692}$$

Mean forecast and Forecast error

$$E\left(y_{T+2}\right) = \widehat{y}_{T+2} = \delta + \theta_1 y_{t+1} \tag{1693}$$

$$\begin{aligned} \widehat{e}_{T+2} &= \left(y_{T+2} - \widehat{y}_{T+2}\right) = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1} - \delta - \theta_1 \widehat{y}_{T+1} \\ &= \theta_1\left(y_{t+1} - \widehat{y}_{T+1}\right) + e_{T+2} + \alpha_1 e_{T+1} = \left(\theta_1 + \alpha_1\right) e_{T+1} + e_{T+2} \end{aligned} \tag{1694}$$

Variance of Forecast error

$$var\left(\widehat{e}_{T+1}\right) = var\left[\left(\theta_1 + \alpha_1\right) e_{T+1} + e_{T+2}\right] = var\left(e_{T+1}\right) = \sigma_e^2\left[\left(\theta_1 + \alpha_1\right)^2 + 1\right] \tag{1695}$$

h=3 period ahead Forecast

$$y_{T+2} = \delta + \theta_1 y_{t+2} + e_{T+3} + \alpha_1 e_{T+2} \tag{1696}$$

Mean forecast

$$E\left(y_{T+3}\right) = \widehat{y}_{T+3} = \delta + \theta_1 \widehat{y}_{t+2} \tag{1697}$$

Forecast error and Variance of Forecast error

$$\begin{aligned}
\widehat{e}_{T+3} &= \left(y_{T+3} - \widehat{y}_{T+3}\right) = \delta + \theta_1 y_{t+2} + e_{T+3} + \alpha_1 e_{T+2} - \delta - \theta_1 \widehat{y}_{T+2} \\
&= \theta_1 \left(y_{t+2} - \widehat{y}_{T+2}\right) + e_{T+3} + \alpha_1 e_{T+2} \\
&= e_{T+3} + \alpha_1 e_{T+2} + \left(\theta_1 + \alpha_1\right) e_{T+2} + e_{T+2}
\end{aligned} \tag{1698}$$

$$\begin{aligned}
var\left(\widehat{e}_{T+3}\right) &= var\left[e_{T+3} + \alpha_1 e_{T+2} + \left(\theta_1 + \alpha_1\right) e_{T+2} + e_{T+2}\right] \\
&= \sigma_e^2 \left[1 + \left(1 + \alpha_1\right)^2 + \left(\theta_1 + \alpha_1\right)^2\right]
\end{aligned} \tag{1699}$$

13.2.4 Statespace form and Kalman Filter

Kalmon filter (KF) is a set of vector and matrix recursions.

It minimises errors in time series model in state space form (SSF) and the OLS in ordinary regression.

It consists of one step ahead prediction of observations and state vectors and corresponding mean square error

Computes the likelihood function using the onestep ahead prediction error decomposition.

It smooths series using Kalman gain.

Measurement equation

$$y_t = Z_t \alpha_t + X_t b + G_t u_t \tag{1700}$$

Transition equation

$$\alpha_{t+1} = T_t \alpha_t + W_t b + H_t u_t \tag{1701}$$

initial condition

$$\alpha_1 = T_t \alpha_t + W_0 b_t + H_0 u_0 \tag{1702}$$

G_t and H_t are error system matrices; $u_t \sim NID\left(0, \sigma_{u_t}^2\right)$ $b = c + B\delta \quad \delta \sim N\left(\mu, \sigma^2 \Lambda\right)$

Runn kalman_run. with kalmanf.m in MATLAB for understanding how a Kalman filter algorithm works in the real world (thanks Santosh Bhattarai for showing me this).

13.2.5 Brownian Motion

Consider a random walk

$$y_t = y_{t-1} + \varepsilon_t \qquad \varepsilon_t \ \tilde{}\ \text{iid } N(0,1) \tag{1703}$$

$$y_t = \epsilon_1 + \epsilon_2 + \epsilon_3 + \epsilon_4 + \epsilon_5 + ... + \epsilon_t \tag{1704}$$

change in value of y_t between periods t and s

$$\begin{aligned} y_t - y_s &= \epsilon_{t+1} + \epsilon_{t+2} + \epsilon_{t+3} + \epsilon_{t+4} + \epsilon_{t+5} + ... + \epsilon_s \\ &\ \tilde{}\ N(0,(s-t)) \end{aligned} \tag{1705}$$

Divide the error $y_t - y_{t-1} = \varepsilon_t$

$$\varepsilon_t = e_{1t} + e_{2t} \tag{1706}$$

Now divite the time interval in infinitely small sections

$$\varepsilon_t = e_{1t} + e_{2t} + ... + e_{Nt} \quad \tilde{}N\left(0,\frac{1}{N}\right) \tag{1707}$$

when N $\longrightarrow \infty$ it is called a Brownian motion, $W(t)$. This is continous stochastic function and has folliwng properties

- $W(0). = 0$
- For any dates $0 \leq t_1 \leq t_2 \leq ... \leq t_k \leq 1$ the changes
- $[W(t_2) - W(t_1)], [W(t_3) - W(t_2)], ..., [W(t_k) - W(t_{k-1})]$ are independent multivariate Gaussian with $[W(s) - W(t)]\ \tilde{}N(0,(s-t))$
- Any realisation of $W(t)$ is continous in t with probability 1.

Other continous time process can be generated from standard Browning motions, as:

$$Z(t) = \sigma W(t) \sim N\left(0, \sigma^2 t\right) \tag{1708}$$

Brownian Motion: Application
stochastic volatility of stock prices with diffusion $\sqrt{v_t}$

$$d\ln S_t = (\mu_{st} + \eta_{st} v_t)\, dt + \sqrt{v_t} dW_t \tag{1709}$$

volatility

$$d\nu_t = k_v(\overline{\nu} - v_t)\, dt + \sigma_v \sqrt{v_t} dW_t \tag{1710}$$

$$d\ln \nu_t = (\overline{\nu} - \kappa \ln v_t) + \sigma_v dW_t \tag{1711}$$

Brownian Motion: Markov Chain Monte Carlo stochastic volatility of stock prices with diffusion $\sqrt{v_t}$

$$d\ln S_t = \sqrt{v_t}dW_t \tag{1712}$$

$$d\ln \nu_t = (\overline{\nu} - \kappa \ln v_t) + \sigma_v dW_t \tag{1713}$$

MCMC uses Bayesian inference to combine a prior distribution over unknown distribution over the unknown parameter vector $p(\Theta, X/Y)$ with the conditional density of the state vector $p(X/Y)$ to obtain joint posterior distribution of parameters $p(\Theta/X, Y)$, $p(X/\Theta, Y)$ $p(X, \Theta/Y)$. (See Singleton)

let p log of stock price follow a standard Brownian motion

$$dp(t) = \mu dt + \sigma dB(t) \tag{1714}$$

Moments of Brownian motion

$E(dB) = 0$; $var(dB) = dt$; $E((dB)(dB)) = dt$; $E((dB)(dB)) = o(dt)$; $E((dB)(dt)) = 0$; $var((dB)(dt)) = o(dt)$;

$$\begin{aligned} (dp)^2 &= (\mu dt + \sigma dB)^2 = \mu^2 dt + \sigma^2 (dB)^2 + 2\mu\sigma (dB)(dt) \\ &= \sigma^2 dt \qquad \text{since } \mu = 0 \end{aligned} \tag{1715}$$

$dp(t)$ is a random variable but not $(dp)^2$.

$$df(p,t) = \frac{\partial f}{\partial p}dp + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial p^2}(dp)^2 \tag{1716}$$

See Campbell et.al (1997).

Ito's lemma for Geometric Brownian Motion

let P stock price follow a geometric Brownian motion

$$P(t) = e^{p(t)} \tag{1717}$$

$$\begin{aligned} dP &= \frac{\partial P}{\partial p}dp + \frac{1}{2}\frac{\partial^2 P}{\partial p^2}(dp)^2 \qquad (1718) \\ &= e^{p(t)}dp + \frac{1}{2}e^{p(t)}(dp)^2 \\ &= P(\mu dt + \sigma dB(t)) + \frac{1}{2}P\left(\sigma^2 dt\ \right) \\ dP &= \left(\mu + \frac{1}{2}\sigma^2\right)Pdt\ + \sigma dPB(t) \end{aligned}$$

$$\frac{dP}{P} = \left(\mu + \frac{1}{2}\sigma^2\right)dt\ + \sigma dB(t) \tag{1719}$$

Instantaneous price change $\frac{dP}{P}$ behaves like a Brownian motion.

13.2.6 Local Linear Trend Model

It includes stochastic trend μ_t plus the noise η_t .

$$y_t = \mu_t + \eta_t \tag{1720}$$

$$\mu_t = \mu_{t-1} + a_t + \epsilon_t \tag{1721}$$

$$a_t = a_{t-1} + \delta_t \tag{1722}$$

here η_t ϵ_t and δ_t are white noises. Change in the random walk $\Delta\mu_t = \mu_t - \mu_{t-1}$ is the itself a random walk plus noise process.

$$a_t = a_0 + \sum_{t=0}^{t} \delta_i \tag{1723}$$

$$\mu_t = \mu_{t-1} + a_0 + \sum_{t=0}^{t} \delta_i + \epsilon_t \tag{1724}$$

$$\mu_t = \mu_0 + \sum_{t=0}^{t} \varepsilon_i + t\,(a_0 + \delta_1) + \delta_2\,(t-1) + \delta_3\,(t-2) + .. + \delta_t \tag{1725}$$

General solution of the model:

$$y_t = y_0 + (\eta_t - \eta_0) + \sum_{t=0}^{t} \varepsilon_i + t\,(a_0 + \delta_1) + \delta_2\,(t-1) + \delta_3\,(t-2) + .. + \delta_t \tag{1726}$$

$$\mu_t = \mu_{t-1} + a_0 + \sum_{t=0}^{t} \delta_i + \epsilon_t \tag{1727}$$

$$\mu_t = \mu_0 + \sum_{t=0}^{t} \varepsilon_i + t\,(a_0 + \delta_1) + \delta_2\,(t-1) + \delta_3\,(t-2) + .. + \delta_t \tag{1728}$$

General solution of the model:

$$y_t = y_0 + (\eta_t - \eta_0) + \sum_{t=0}^{t} \varepsilon_i + t\,(a_0 + \delta_1) + \delta_2\,(t-1) + \delta_3\,(t-2) + .. + \delta_t \tag{1729}$$

13.2.7 Optimal Forecast Model

How can one make optimal mix of two forecasts to reduced MSFE?

(See the compilation of forecasts for the UK made by more than 400 forecasters or the BOEs inflation report).

Ganger and Newbold (1986) had illustrated this with h period ahead forecast $f_{\ i}$ for $(i = 1, 2)$ for forecast of y_t.

Forecast error

$$e_{it} = y_t - f_{it} \tag{1730}$$

$e_{it} \sim N(0, \sigma_i^2)$ $\rho = \frac{cov(e_{1t}e_{2t})}{\sqrt{var(e_{1t})}\sqrt{var(e_{2t})}}$

Composite forecast for $0 < \lambda < 1$

$$f_{ct} = (1 - \lambda) f_{1t} + \lambda f_{2t} \tag{1731}$$

Composite error

$$e_{ct} = y_t - f_{ct} = (1 - \lambda) e_{1t} + \lambda e_{2t} \tag{1732}$$

The mean of the composite error

$$E(e_{ct}) = E(y_t - f_{ct}) = (1 - \lambda) E(e_{1t}) + \lambda E(e_{2t}) = 0 \tag{1733}$$

$$\begin{aligned} Var(e_{ct}) &= E\left[(y_t - f_{ct})\right]^2 = \left[(1 - \lambda) E(e_{1t}) + \lambda E(e_{2t})\right]^2 \\ &= (1 - \lambda)^2 \sigma_1^2 + \lambda^2 \sigma_2^2 + 2(1 - \lambda) \lambda \rho \sigma_1 \sigma_2 \end{aligned} \tag{1734}$$

The optimal λ the weight to be attached to a certain forecast is chosen by minimising this variance:

$$\frac{\partial Var(e_{ct})}{\partial \lambda} = -2(1 - \lambda) \sigma_1^2 + 2\lambda \sigma_2^2 + 2\rho\sigma_1\sigma_2 - 4\rho\lambda\sigma_1\sigma_2 = 0 \tag{1735}$$

$$\lambda_{opt} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{1736}$$

Unknown values of σ_1^2,σ_2^2, 2ρ are estimated from sample information using the residuals as:

$$\widehat{\lambda}_{opt} = \frac{\sum\limits_{t=1}^{T} e_{1t}^2 - \sum\limits_{t=1}^{T} e_{1t}e_{2t}}{\sum\limits_{t=1}^{T} e_{1t}^2 + \sum\limits_{t=1}^{T} e_{2t}^2 - 2\sum\limits_{t=1}^{T} e_{1t}e_{2t}} \tag{1737}$$

13.3 Minimising the Square Errors (MSE): Signal extraction

Minimising the Square Errors (MSE): Signal extraction

Let the estimated function be

$$y^* = a + b(x - \mu_x) \tag{1738}$$

$$E(y - y^*)^2 = E[y - a - b(x - \mu_x)]^2 \tag{1739}$$

$$E(y - y^*)^2 = E\left[\begin{array}{c} y^2 + a^2 + b^2(x - \mu_x)^2 \\ -2ya + 2ab(x - \mu_x) - 2yb(x - \mu_x) \end{array}\right] \tag{1740}$$

$E(x - \mu_x) = 0$; $E(y) = 0$; $E(x - \mu_x)^2 = \sigma_x^2$; $E(yx) - \mu_x\mu_y = cov(x, y) = \sigma_{xy}$

$$E(y - y^*)^2 = MSE = Ey^2 + a^2 + b^2\sigma_x^2 - 2a\mu_y - 2b\sigma_{xy} \tag{1741}$$

minimise MSE wrt to a and b

$2a - 2\mu_y = 0 ==> \mu_y = a$ and $\frac{\partial MSE}{\partial b} = 2b\sigma_x^2 - 2\sigma_{xy} = 0$; $b = \frac{\sigma_{xy}}{\sigma_x^2}$

Minimising the Square Errors (MSE): Signal extraction

Optimal prediction is

$$y^* = a + b(x - \mu_x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}x - \frac{\sigma_{xy}}{\sigma_x^2}\mu_x \tag{1742}$$

Forecast is unbiased

$$Ey^* = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}\mu_x - \frac{\sigma_{xy}}{\sigma_x^2}\mu_x = \mu_y \tag{1743}$$

Signal extraction (series $\{y_t\}$ includes two components $\{\varepsilon_t\}$ and $\{\eta_t\}$ permanent and transitory shocks; where $E\{\varepsilon_t\} = 0; E\{\eta_t\} = 0; E(\varepsilon_t\eta_t) = 0; E\left\{\varepsilon_t^2\right\} = \sigma^2; \left\{\eta_t^2\right\} = \sigma_\eta^2$

$$y_t = \epsilon_t + \eta_t \tag{1744}$$

$$\epsilon_t^* = a + by_t \tag{1745}$$

Minimising the Square Errors (MSE): Signal extraction

$$E(\epsilon_t - \epsilon_t^*)^2 = E(\epsilon_t - by_t)^2 = E(\epsilon_t - b(\epsilon_t + \eta_t))^2 \tag{1746}$$

$$\begin{aligned} MSE &= E(\epsilon_t - \epsilon_t^*)^2 = E(\epsilon_t - by_t)^2 = E[(1 - b)\epsilon_t - b\eta_t]^2 \\ &= (1 - b)^2\sigma^2 + b^2\sigma_\eta^2 \end{aligned} \tag{1747}$$

$$\frac{\partial MSE}{\partial b} = -2(1 - b)\sigma^2 + 2b\sigma_\eta^2 = 0 \Rightarrow b = \frac{\sigma^2}{\sigma^2 + \sigma_\eta^2} \tag{1748}$$

Partitioning parameter b is determined by the respective variances. Remaining signal is extracted as a residual.

Minimising the Square Errors (MSE): Signal extraction

$$\eta_t^* = y_t - \epsilon_t^* \tag{1749}$$

Hodrik-Prescot decomposition of trend and stationary components:

$$HP = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mu_t)^2 + \frac{\lambda}{T}\sum_{t=1}^{T}\left(\mu_{t+1} - \mu_t\right)\left(\mu_t - \mu_{t-1}\right) \tag{1750}$$

Problem is to choose sequence of μ_t to minimiseHP; if $\lambda = 0$ then $y_t = \mu_t$ implying that y_t is trend in itself. Larger λ closer HP is to the linear trend $\Delta\mu_{t+1}$ and $\Delta\mu_t$ become smaller and smaller. with $\lambda \to \infty$ the HP approaches linear trend.

13.3.1 Nelson-Beberidge decomposition

Forecast function in Nelson-Beberidge decomposition

$$y_{t+s} = a_0 s + y_t + \sum_{i=0}^{s}\varepsilon_{t+i} + \beta_1\sum_{i=1}^{s}\varepsilon_{t+i-1} + \beta_2\sum_{i=1}^{s}\varepsilon_{t+i-2} \tag{1751}$$

This is derived as following:

Random walk with a drift

$$y_t = y_{t-1} + a_0 + \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} \tag{1752}$$

Define $e_t = \epsilon_t + \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2}$

$$y_2 = y_1 + a_0 + e_t \tag{1753}$$

$$y_t = y_0 + ta_0 + \sum_{t=0}^{t}e_i \tag{1754}$$

Forecast function

$$y_{t+s} = y_0 + (t+s)\,a_0 + \sum_{i=1}^{t+s}e_i \tag{1755}$$

Nelson-Beberidge decomposition

$$y_{t+s} = y_t + sa_0 + \sum_{i=1}^{s}e_{t+i} \tag{1756}$$

$$\sum_{t=0}^{t}e_i = \sum_{t=0}^{t}\epsilon_t + \beta_1\sum_{t=0}^{t}\epsilon_{t-1} + \beta_2\sum_{t=0}^{t}\epsilon_{t-2} \tag{1757}$$

13.3.2 Box-Jenkins Approach to Forecasting

Main use of ARMA(p,q) model is in forecasting values of $\{y_t\}$ on the basis of assumed DGP and the white noises, (ϵ_t). For instance consider AR(1) process

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t \tag{1758}$$

By iterating forward one period ahead forecast would be

$$y_{t+1} = a_0 + a_1 y_t + \epsilon_{t+1} \tag{1759}$$

$$y_{t+2} = a_0 + a_1 y_{t+1} + \epsilon_{t+2} \tag{1760}$$

here a_0, a_1 are known and $\epsilon_t \sim N(0, \sigma_\varepsilon^2)$ and information set $\Omega_t \left\{y_{t+j}/y_{t-1}, y_{t-2}, ...y_{t-n}, \epsilon_t, \epsilon_{t-1}, ...\epsilon_{t-j}\right\}$
The conditional forecast to y_{t+1} made at time t is expressed as a forecast function:

$$E_t y_{t+1} = a_0 + a_1 y_t \tag{1761}$$

Similarly conditional forecast to y_{t+2} made at time t is

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1} = a_0 + a_1 \left(a_0 + a_1 y_t\right) = a_0 + a_0 a_1 + a_1^2 y_t \tag{1762}$$

Similarly the three period ahead forecast is

$$\begin{aligned} E_t y_{t+3} &= a_0 + a_1 E_t y_{t+2} = a_0 + a_1 \left(a_0 + a_0 a_1 + a_1^2 y_t\right) \\ &= a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 y_t \end{aligned} \tag{1763}$$

In general the j period ahead forecast

$$E_t y_{t+j} = a_0 + a_1 E_t y_{t+j-1} = a_0 + a_0 a_1 + a_0 a_1^2 + ... + a_0 a_1^{j-1} + a_1^j y_t \tag{1764}$$

Forecast converges to $\frac{a_0}{1-a_1}$ if $|a_1| < 1$.

$$\underset{j\to\infty}{E_t y_{t+j}} = \frac{a_0}{1-a_1} \tag{1765}$$

Forecast error (unforecastable part of future y_t)
One period ahead forecast error

$$E_t y_{t+1} = a_0 + a_1 y_t \tag{1766}$$

$$fe_t(1) = y_{t+1} - E_t y_{t+1} = \epsilon_{t+1} \tag{1767}$$

$$E\left[fe_t(1)\right] = 0; Var\left[fe_t(1)\right] = \sigma^2 \tag{1768}$$

two period ahead forecast error

$$\begin{aligned} y_{t+2} &= a_0 + a_1 y_{t+1} + \epsilon_{t+2} = a_0 + a_1 \left(a_0 + a_1 y_t + \epsilon_{t+1}\right) + \epsilon_{t+2} \\ &= a_0 + a_0 a_1 + a_1^2 y_t + a_1 \epsilon_{t+1} + \epsilon_{t+2} \end{aligned} \tag{1769}$$

$$fe_t(2) = y_{t+2} - E_t y_{t+2} = \epsilon_{t+2} + a_1 \epsilon_{t+1} \tag{1770}$$

$$E[fe_t(2)] = E(\epsilon_{t+2}) + a_1 E(\epsilon_{t+1}) = 0 \tag{1771}$$

$$var[fe_t(2)] = \sigma^2 + a_1^2 \sigma^2 = \sigma^2 \left(1 + a_1^2\right) \tag{1772}$$

j period ahead forecast error

$$fe_t(j) = y_{t+j} - E_t y_{t+j} = \epsilon_{t+j} + a_1 \epsilon_{t+j-1} + a_1^2 \epsilon_{t+j-2} + ... + a_1^{j-1} \epsilon_{t-1} \tag{1773}$$

$E[fe_t(j)] = 0$

$$var[fe_t(j)] = \sigma^2 \left(1 + a_1^2 + a_1^4 + a_1^6 + ... + a_1^{2(j-1)}\right) \tag{1774}$$

It is obvious from above examples that the variance of forecast rises as the horizon of forecast rises. There is more confidence in the short run forecast than in the long run forecast. With $|a_1| < 1$ the forecast variance converges to $\frac{\sigma^2}{1-a_1^2}$.

Confidence interval of forecast (assuming normality of the errors,$\epsilon_t \sim N(0, \sigma_\varepsilon^2)$)

For one period ahead forecast:

$$95\% \text{ confidence interval for } E_t y_{t+1} = a_0 + a_1 y_t \pm 1.96\sigma \tag{1775}$$

For two period ahead forecast:

$$95\% \text{ confidence interval for } E_t y_{t+1} = a_0 + a_1 y_t + a_1^2 y_t \pm 1.96\sigma^2 \left(1 + a_1^2\right) \tag{1776}$$

VAR Based Forecasting

$$X_t = A_0 + A_1 X_{t-1} + e_t \tag{1777}$$

from successive iteration this reduces to

$$E_t X_{t+n} = \left(I + A_1 + A_1^2 + A_1^3 + ... + A_1^{n-1}\right) A_0 + A_1^n X_t + e_t \tag{1778}$$

Forecast error is given by

$$\left(e_{t+n} + A_1 e_{t+n-1} + A_1^2 e_{t+n-2} + ... + A_1^{n-1} e_{t+1}\right) \tag{1779}$$

$$E_t X_{t+n} = \mu + \sum_{i=0}^{n-1} \phi_i(i) \epsilon_{t+n-i} \tag{1780}$$

$$X_{t+n} - E_t X_{t+n} = \sum_{i=0}^{n-1} \phi_i(i) \epsilon_{t+n-i} \tag{1781}$$

Taking only one equation

$$
\begin{aligned}
y_{t+n} - E_t y_{t+n} &= \phi_{11}(0)\epsilon_{yt+n} + \phi_{11}(1)\epsilon_{yt+n-1} + ... + \phi_{11}(n-1)\epsilon_{yt+1} \\
&\quad + \phi_{12}(0)\epsilon_{zt+n} + \phi_{12}(1)\epsilon_{zt+n-1} + ... + \phi_{12}(n-1)\epsilon_{zt+1} \qquad (1782)
\end{aligned}
$$

Variance of n-step ahead forecast error is

$$
\begin{aligned}
\sigma(n)_y^2 &= \sigma_y^2\left[\phi_{11}(0) + \phi_{11}(1) + ... + \phi_{11}(n-1)\right] \\
&\quad + \sigma_z^2\left[\phi_{12}(0) + \phi_{12}(1) + ... + \phi_{12}(n-1)\right] \qquad (1783)
\end{aligned}
$$

Variance decomposition in terms of variances of shocks ϵ_{yt} and ϵ_{zt}.

$$
\begin{aligned}
1 &= \frac{\sigma_y^2\left[\phi_{11}(0) + \phi_{11}(1) + ... + \phi_{11}(n-1)\right]}{\sigma(n)_y^2} \\
&\quad + \frac{\sigma_z^2\left[\phi_{12}(0) + \phi_{12}(1) + ... + \phi_{12}(n-1)\right]}{\sigma(n)_y^2} \qquad (1784)
\end{aligned}
$$

Thus the variance decomposition is finding the proportion of variance explained by variables its own shock(ϵ_{yt}) versus the variance explained by shock of the over variable (ϵ_{yt}).

13.4 Clement and Hendry (2000) Explanation of Forecast Failures

Clement and Hendry (2000) Explanation of Forecast Failures

They have the following example to illustrate this.

$$
y_t = \phi + \Pi y_{t-1} + \epsilon_t \qquad (1785)
$$

where $\epsilon_t \sim N(0, \sigma_y^2)$

$$
y_t = (I - \Pi)^{-1}\phi = \varphi \qquad (1786)
$$

Deviation from the forecast

$$
y_t - \varphi = \Pi\left(y_{t-1} - \varphi\right) + \epsilon_t \qquad (1787)
$$

$h-$step ahead forecast

$$
\widehat{y}_{T+h} - \widehat{\varphi} = \widehat{\Pi}\left(\widehat{y}_{T+h-1} - \widehat{\varphi}\right) = \widehat{\Pi}^h\left(\widehat{y}_{T+h-1} - \widehat{\varphi}\right) \qquad (1788)
$$

where $\varphi = (I - \Pi)^{-1}\phi$

After forecast at time t is made $\phi : \Pi$ are allowed change to $\phi^* : \Pi^*$

$$
y_{T+1} = \phi^* + \Pi^* y_{T+h-1} + \epsilon_{T+1} \qquad (1789)
$$

$\phi^* = \left(I_n - \Pi^*\right)^{-1}\varphi^*$

$$
\begin{aligned}
y_{T+h} - \varphi^* &= \Pi^*\left(y_{T+h-1} - \varphi^*\right) + \epsilon_{T+h} = \left(\Pi^*\right)^h\left(y_{T+h-1} - \varphi^*\right) \\
&\quad + \sum_{i=0}^{h-1}\left(\Pi^*\right)^h \epsilon_{T+h-i} \qquad (1790)
\end{aligned}
$$

$$\widehat{\varepsilon}_{T+h/T} = y_{T+h} - \widehat{y}_{T+h} = \varphi^* - \widehat{\varphi} + (\Pi^*)^h (y_{T+h} - \varphi^*) - \widehat{\Pi}(\widehat{y}_{T+h} - \widehat{\varphi}) + \sum_{i=0}^{h-1} (\Pi^*)^h \epsilon_{T+h-i} \quad (1791)$$

These errors are decomposed in (i) slope change (ii) equilibrium mean change (iii) slope mispecification (iv) equilibrium mean mispecification (v) slope estimation (vi) equilibrium mean estimation (viii) initial condition uncertainty and (viiii) error accumulation.

Do you remember difference and differential equations ? These models give time path of variables.

$$y_t = \left(y_0 - \frac{b}{a}\right) e^{-at} + \frac{b}{a} \quad (1792)$$

$$y_t = \left[y_0 - \frac{B}{A}\right] e^{-At} + \frac{B}{A} = \left[y_0 - \frac{\alpha\left(a + \frac{n\overline{M}}{M} + I + G\right)}{\alpha\left(1 - b + \frac{nk}{h}\right)}\right] e^{-\alpha\left(1-b+\frac{nk}{h}\right)t} + \frac{\alpha\left(a + \frac{n\overline{M}}{M} + I + G\right)}{\alpha\left(1 - b + \frac{nk}{h}\right)} \quad (1793)$$

$$y_t = \left(y_0 - \frac{\alpha\ +\gamma}{\beta + \delta}\right) e^{-k(\beta+\delta)t} + \frac{\alpha\ +\gamma}{\beta + \delta} \quad (1794)$$

$$P_t = P_c + P_p = e^{-\frac{m}{2n}t}\left[A_5 \cos(vt) + A_6 \,\mathrm{Si}\,n\,(vt)\right] + \frac{\alpha\ +\gamma}{\beta + \delta} \quad (1795)$$

$$P_t = A_1 e^{r_1 t} + A_2 e^{r_2 t} + 4 = e^{6t} + e^{-2t} + 4 \quad (1796)$$

13.4.1 Predition

$$Y_0 = \beta_1 + \beta_2 X_i + \varepsilon_0 \quad (1797)$$

$$\varepsilon_0 \sim N\left(0, \sigma^2\right) \quad (1798)$$

Mean prediction:

$$\widehat{Y}_0 = \widehat{\beta}_1 + \widehat{\beta}_2 X_i \quad (1799)$$

Prediction error

$$f = Y_0 - \widehat{Y}_0 = \beta_1 + \beta_2 X_i + \varepsilon_0\ - \widehat{\beta}_1 - \widehat{\beta}_2 X_i \quad (1800)$$

Mean of prediction error

$$E(f) = E\left(\beta_1 + \beta_2 X_i + \varepsilon_0\ - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\ \right) = 0 \quad (1801)$$

Predictor is ubiased.
Variance of Prediction error:

$$var\left(\widehat{\beta}_1\right) = \left[1 + \frac{1}{N} + \frac{\left(x_0 - \overline{x}\right)^2}{\sum\left(x_0 - \overline{x}\right)_i^2}\right]\widehat{\sigma}^2 \tag{1802}$$

Proof

$$Y_0 = \widehat{Y}_0 + \widehat{\varepsilon}_0 \tag{1803}$$

$$var\left(Y_0\right) = var\left(\widehat{Y}_0\right) + var\left(\widehat{\varepsilon}_0\right) \tag{1804}$$

$$var\left(\widehat{Y}_0\right) = var\left(\widehat{\beta}_1 + \widehat{\beta}_2 X_0\right) = var\left(\widehat{\beta}_1\right) + X_0^2 var\left(\widehat{\beta}_2\right) + 2X_0 covar\left(\widehat{\beta}_1\widehat{\beta}_2\right) \tag{1805}$$

Variance of Prediction

$$\begin{aligned} var\left(\widehat{Y}_0\right) &= \frac{\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 + X_0^2\frac{\sum\left(X_i - \overline{X}\right)}{\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 \\ &\quad +2X_0\left(-\overline{X}\frac{1}{\sum\left(X_i - \overline{X}\right)^2}\right)\widehat{\sigma}^2 \end{aligned} \tag{1806}$$

add and subtract $\frac{N\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2$

$$\begin{aligned} var\left(\widehat{Y}_0\right) &= \frac{\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 - \frac{N\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 + X_0^2\frac{\sum\left(X_i - \overline{X}\right)}{\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 \\ &\quad +2X_0\left(-\overline{X}\frac{1}{\sum\left(X_i - \overline{X}\right)^2}\right)\widehat{\sigma}^2 + \frac{N\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2}\widehat{\sigma}^2 \end{aligned} \tag{1807}$$

Variance of forecast: taking common elements out

$$var\left(\widehat{Y}_0\right) = \widehat{\sigma}^2\left[\begin{array}{c} \frac{\sum\left(X_i - \overline{X}\right)^2 - N\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2} \\ +\frac{X_0^2 - 2X_0\overline{X} + \sum\left(X_i - \overline{X}\right)^2}{\sum\left(X_i - \overline{X}\right)^2} \end{array}\right] \tag{1808}$$

$$var\left(\widehat{Y}_0\right) = \widehat{\sigma}^2\left[\frac{\sum\left(X_i - \overline{X}\right)^2}{N\sum\left(X_i - \overline{X}\right)^2} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum\left(X_i - \overline{X}\right)^2}\right] \tag{1809}$$

$$var\left(\widehat{Y}_0\right) = \widehat{\sigma}^2\left[\frac{1}{N} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum\left(X_i - \overline{X}\right)^2}\right] \tag{1810}$$

$$var\left(f\right)=var\left(\widehat{Y}_0\right)+var\left(\widehat{\varepsilon}_0\right) \tag{1811}$$

$$var\left(f\right)=\widehat{\sigma}^2\left[\frac{1}{N}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right]+\widehat{\sigma}^2 \tag{1812}$$

$$var\left(f\right)=\widehat{\sigma}^2\left[1+\frac{1}{N}+\frac{\left(X_0-\overline{X}\right)^2}{\sum\left(X_i-\overline{X}\right)^2}\right] \tag{1813}$$

t-test for variance of forecast

$$t_f=\frac{Y_0-\widehat{Y}_0}{SE\left(f\right)}\sim t_{N-2} \tag{1814}$$

Standard error of forecast

$$SE\left(f\right)=\sqrt{var\left(f\right)} \tag{1815}$$

Confidence interval of forecast

$$\Pr\left[-t_c\leq\frac{Y_0-\widehat{Y}_0}{SE\left(f\right)}\leq t_c\right]=\left(1-\alpha\right) \tag{1816}$$

$$\Pr\left[\widehat{Y}_0-t_cSE\left(f\right)\leq Y_0\leq\widehat{Y}_0+t_cSE\left(f\right)\right]=\left(1-\alpha\right) \tag{1817}$$

13.5 Non parametric estimation

- Hardle, W. and Linton, O. (1994) Applied nonparametric methods, in (D. F. McFadden and R. F. Engle III, eds.) The Handbooko f EconometricsVol. IV, Amsterdam: North-Holland.
- Campbell J. Y., A. W. Lo and A C MacKindlay (1997) The Econometrics of Financial Markets, Princeton.

These are smoothing estimators and main types include:

- Kernel regression
- Orthogonal series expansion
- Projection pursuit
- Nearest-neighour estimator
- Average deviation estimators
- Splines
- Artificial neural networks.

13.5.1 Kernel regression

$$Y_t = m\left(x_t\right) + \varepsilon_t; \qquad t = 1, .., ., \ T \tag{1818}$$

Here $m\left(x_t\right)$ is arbitrarily fixed but given by a underlying non-linear function. For a particular data point $X_{t=0} = x_0$ there are repeated independent observations of $Y_{t=0} = \{y_1, y_2, ...y_n\}$. Then by the law of large numbers

$$\widehat{m}\left(x_0\right) = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} y_i \left[m\left(x_0\right) + \varepsilon_t^i\right] = m\left(x_0\right) + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_t^i = m\left(x_0\right) \tag{1819}$$

More generally

$$\widehat{m}\left(x_0\right) = \frac{1}{T}\sum_{i=1}^{T}\omega_{t,T(x)} y_t \tag{1820}$$

with $\omega_{t,T(x)}$ weights assigned according to the distance.

In the Kernel regression optimal weights $\omega_{t,T(x)}$ are constructed from a kernel probability density function: $\int k(u)du = 1$

Nadaraya-Watson kernel estimator with bandwidth parameter h

$$\widehat{m}\left(x_0\right) = \frac{1}{T}\sum_{i=1}^{T}\omega_{t,T(x)} y_t = \frac{\sum_{i=1}^{T} K_h\left(x - X_t\right) Y_t}{\sum_{i=1}^{T} K_h\left(x - X_t\right)} \tag{1821}$$

An example of kernel regression:

$$Y_t = \sin\left(X_t\right) + 0.7\varepsilon_t \tag{1822}$$

Average derivative estimator can be of the form of the direct, indirect and slope estimators; with $k$number of X variables

$$Y_t = m\left(X_t\beta\right) + \varepsilon_t; \qquad E\left(\varepsilon_T | X_t\right) = 0 \tag{1823}$$

Indirect slope estimation (ISE) is estimated using an instrument H

$$\beta_{ISE} = \left(H'X\right)^{-1} H'Y \tag{1824}$$

13.5.2 Artificial Neural Network

It is a learning network consisting of multilayer perceptons (MLP), radial basis function (RBF), hidden layers between the input layer and the output layers. A standard representation is :

$$Y = q\left(\sum_{j=1}^{J}\beta_j X_j - \mu\right) \tag{1825}$$

where β_j is the connection strength.

$$g(u) = \int_{0..... \; if \;\; u \leqslant 0}^{1 \qquad if \;\; u \geqslant 0} \tag{1826}$$

If the sum $\left(\sum_{j=1}^{J} \beta_j X_j \right)$ exceeds the threshold (μ) then the artificial neuron is switched on or activated. Similarly the projection pursuit regression (PPR) aims at analysing high dimensional dataset by looking at low dimensional projections

$$m\left(x_0\right) = \alpha_0 + \sum_{k=1}^{K} \alpha_k m_k \left(\beta_t' X_t\right) \tag{1827}$$

and the radial basis function (RBF) aims to minimise the objective functional of the form

$$v\left(m\right) = \sum_{i=1}^{T} \left(\left\| \widehat{Y}_t - m\left(x_t\right) \right\|^2 + \lambda \left\| Dm\left(x_t\right) \right\|^2 \right) \tag{1828}$$

where $\left[\widehat{Y}_t - m\left(x_t\right)\right]$ measures the distance between the $m\left(x_t\right)$ and the observation Y_t and $Dm\left(x_t\right)$ is penalty function decreasing with the smoothness of $m\left(x_t\right)$ and the λ controls the trade-off between smoothness and fit.

13.5.3 State price densities

Economic application of non-parametric techniques is in deriving state price densities of the Arrow-Debreau (1064) securities where the equilibrium price p_t at time t of a security of single liquidating payoff $Y\left(C_T\right)$ at data T is given by:

$$P_t = E_t\left[Y\left(C_T\right) M_{t,T}\right]; \quad M_{t,T} = \frac{\delta^{T-1} U'(C_T)}{U'(C_t)} \tag{1829}$$

$$P_t = e^{-r_{t,T(T-1)}} \int Y\left(C_T\right) f^*\left(C_T, t, T\right) dc_T; \quad f^* = \frac{M_{t,T} f_t(C_T)}{\int M_{t,T} f_t(C_t) dc_t} \tag{1830}$$

As explained in Campbell Lo and MacKinlay (1997, page 509) the second derivative of the call pricing function G_t with respect to strike price X_t must equal the state price density (SPD)

$$\frac{\partial^2 G_t}{\partial X^2} = e^{-r_{t,T(T-1)}} \; f^* \tag{1831}$$

Parametric formula of call pricing of Black and Scholes (1973) in Brownian motion are good when assumptions are true but the non-parametric estimation of call option pricing is more robust and based on fewer assumptions (Alt Sahalia an Lo (1996))

Using a multivariate kernel

$$K_h\left(P, X, \tau, r_t\right) = k_{h_p}\left(P\right) k_{h_x}\left(X\right) k_{h_\tau}\left(\tau\right) k_{h_t}\left(r_\tau\right) \tag{1832}$$

$$\widehat{G}\left(P, X, \tau, r_t\right)=\frac{\sum_{i=1}^{T} k_{h_p}\left(P-P_i\right) k_{h_x}\left(X-X_i\right) k_{h_\tau}\left(\tau-\tau_i\right) k_{h_r}\left(r_\tau-r_{\tau_i}\right) C_i}{\sum_{i=1}^{T} k_{h_p}\left(P-P_i\right) k_{h_x}\left(X-X_i\right) k_{h_\tau}\left(\tau-\tau_i\right) k_{h_r}\left(r_\tau-r_{\tau_i}\right)} \tag{1833}$$

Option's delta and state price densities (SPD) are:

$$\widehat{\Delta}\left(P, X, \tau, r_t\right)=\frac{\partial\left(P, X, \tau, r_t\right)}{\partial P} \tag{1834}$$

$$f^*\left(P_T | P, , \tau, r_t\right)=e^{r_t \tau}\left[\frac{\partial^2\left(P, X, \tau, r_t\right)}{\partial X^2}\right]_{X=P_T} \tag{1835}$$

These delta and SPD are consistent and asymptotically normal as shown by Alt Sahalia an Lo (1996).

These non-linear estimation techniques suffer from "overfitting" or data-snooping problems. Overfitting occurs because of too many parameters relative to the number of data points and data-snooping occurs because the fit is spurious and result of an extensive search procedure. A priori theoretical considerations and good mathematical models of economic behavior can overcome this problem.

14 LASSO and Ridge Regression (Machine Learning)

Least Absolute Selection and Shrinkage Operator (LASSO)

Theories of LASSO and Ridge Regression (Knight and Fu (2000) : a special case of penalized regression or Bridge estimator

$$Min \sum\left(Y_i-X^T \phi\right)^2+\lambda_n|\phi|^\gamma \tag{1836}$$

$\gamma=2 \Longrightarrow$ ridge regression and $\gamma=1 \Longrightarrow$ Lasso regression. LS with restriction.

Authors follow Hierarchical LASSO objectives by Bein et al. (2013)

$$Y=\beta_0+\sum_j^p \beta_j X_j+\frac{1}{2} \sum_{j \neq k}^p \theta_{jk} X_j X_k+U \tag{1837}$$

Least Absolute Selection and Shrinkage Operator (LASSO)

$$Min \sum\left(Y-\beta_0-\sum_j^p \beta_j X_j-\frac{1}{2} \sum_{j \neq k}^p \theta_{jk} X_j X_k+U\right)^2+\lambda \sum_{j=1}^p\left\|\beta_g\right\|+\lambda \sum_{j=k}^p\|\theta\| \tag{1838}$$

Group LASSO objectives by Yuan and Lin (2016)

$$Min \sum\left(Y-\beta_0-\sum_{g=1}^G \beta_g X_j\right)^2+\sum_{g=1}^G\left\|\beta_g\right\| \tag{1839}$$

https://www.youtube.com/results?search_query=lasso+regression
Justification of the research technique

- It is important provide intuition behind the LASSO
- How does it shrink parameters between treating and testing samples.
- How is sample divided between Treating and Testing Samples
- Small bias accepted in treating sample in order to reduce the variance in the overall regression
- Compare with the Ridge regression : while Ridge regression squares slope parameters in restriction and LASSO has absolute deviations
- Rigorous theory of consistency of LASSO is discussed on Chatterjee, A., and S. N. Lahiri (2008)

Strengths and assymptotic properties of LASSO Estimators

- Two main benefits of the Lasso are:

(i) the nature of regularization used in the Lasso leads to sparse solutions, which automatically leads to parsimonious model selection (see Zhao and Yu (2006), Wainwright (2006), Zou (2006)) and (ii) it is computationally feasible (see Efron et. al (2004), Osborne et al. (2000), Fu (1998)), even in high dimensional settings.

- The asymptotic properties of the Lasso was first studied by Knight and

Fu (2000) for the finite dimensional regressions
$\left\|\widehat{\beta}_n - \beta\right\| = 0\left(n^{-\frac{(\alpha-1)}{\alpha}}\right)$ with prob =1 as
$\left\|\widehat{\beta}_{OLS} - \beta\right\| = 0\left(n^{-\frac{(\alpha-1)}{\alpha}}\right)$ with prob =1

- Further literature: particularly on theory of LASSO and Ridge Regression

Belloni, A., Chernozhukov, V., Fernández-Val, I. and Hansen, C., 2017 (with programs and data supplements in Econometrica)
Chatterjee, A., and S. N. Lahiri (2008)
Knight, Keith, and Wenjiang Fu (2000)
Hans, C., 2009; Hans, C., 2010 ;
Hansen, Bruce E. (2016)
Liangjun su, zhentao shi, and peter c. B. Phillips (2016)
Wang, Hansheng, Guodong Li, and Guohua Jiang (2007)
Varian, Hal R. 2014.

References

[1] Belloni, A., Chernozhukov, V., Fernández-Val, I. and Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. Econometrica, 85(1), pp.233-298.

[2] Chatterjee, A., and S. N. Lahiri. "Strong Consistency of Lasso Estimators." Sankhyā: The Indian Journal of Statistics, Series A (2008-) 73, no. 1 (2011): 55-78.

[3] Frank, L.E. and Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics, 35(2), pp.109-135.

[4] Hansen, Bruce E.; The Risk of James-Stein and Lasso Shrinkage Econometric Reviews, 2016, v. 35, iss. 8-10, pp. 1456-70

[5] Hans, C., 2010. Model uncertainty and variable selection in Bayesian lasso regression. Statistics and Computing, 20(2), pp.221-229.

[6] Hans, C., 2009. Bayesian lasso regression. Biometrika, 96(4), pp.835-845.

[7] Knight, Keith, and Wenjiang Fu. "Asymptotics for Lasso-Type Estimators." The Annals of Statistics 28, no. 5 (2000): 1356-378.

[8] Liangjun su, zhentao shi, and peter c. B. Phillips (2016) Identifying latent structures in panel data, Econometrica, Vol. 84, No. 6 (November, 2016), 2215–2264

[9] Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." Journal of Economic Perspectives 28 (2): 3–28

[10] Wang, Hansheng, Guodong Li, and Guohua Jiang. "Robust Regression Shrinkage and Consistent Variable Selection through the LAD-Lasso." Journal of Business & Economic Statistics 25, no. 3 (2007): 347-55.

[11] Zhao, P. and Yu, B., 2006. On model selection consistency of Lasso. Journal of Machine learning research, 7(Nov), pp.2541-2563.

[12] Bien, Jacob, Jonathan Taylor, and Rob Tibshirani, "A Lasso for Hierarchical Interactions." Annals of Statistics. 41(3). (2013) , 1111-1141

[13] Friedman, Jerome, Trevor Hastie, and Rob Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software, Vol. 33(1), January 2010, 1-22.

[14] Mullainathan, Sendhil, and Jann Spiess. "Machine Learning: An Applied Econometric Approach." Journal of Economic Perspectives, 31(2) (2017). 87-106.

[15] Tibshirani, Robert. "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) 58, no. 1 (1996): 267-88. http://www.jstor.org/stable/2346178.

[16] Yuan, M., & Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 68(1), 49-67. Retrieved from http://www.jstor.org/stable/3647556

[17] Simon, Noah, Jerome Friedman, Trevor Hastie, and Robert Tibshirani.C2013. "A Sparse-Group Lasso." Journal of Computational and Graphical Statistics 22 (2): 231–45. https://doi.org/10.1080/10618600.2012.681250.

References

[1] Artis, M. and Marcellino, M. (2001) Fiscal forecasting: The track record of the IMF, OECD and EC. The Econometrics Journal, 4: 20–36.

[2] Bradford D F and H. H. Kelejian (1977) The Value of Information for Crop Forecasting in a Market System: Some Theoretical Issues, Review of Economic Studies (1977) 44 (3): 519-531

[3] Bai J and S. Ng (2008) Forecasting economic time series using targeted predictors Journal of Econometrics, 146, 2, October 2008, 304-317

[4] Blundell R. W.K. Newey and T. Persson (2006) Advances in Economics and Econometrics, Econometric Society Monograph, Vol. II and III, Cambridge University Press.

[5] Campbell J. Y., A. W. Lo and A C MacKindlay (1997) The Econometrics of Financial Markets, Princeton.

[6] Clement MP and DF Hendry ed. (2002) A Companion to Economic Forecasting, Blackwell.

[7] Clements, M. P. and Hendry, D. F. (2002), Modelling methodology and forecast failure. The Econometrics Journal, 5: 319–344.

[8] Clements, M. P. and Krolzig, H.-M. (1998), A Comparison of the Forecast Performance of Markov-switching and Threshold Autoregressive Models of US GNP. The Econometrics Journal, 1: 47–75.

[9] Elliott G, A. Timmermann and I. Komunjer (2005) Estimation and Testing of Forecast Rationality under Flexible Loss Review of Economic Studies (2005) 72 (4): 1107-1125

[10] Ericsson, N. R. and Marquez, J. (1998), A Framework for Economic Forecasting. The Econometrics Journal, 1: 203–227

[11] Engle R.F and M. E. Sokalska (2012) Forecasting intraday volatility in the US equity market. Multiplicative component GARCH, Journal Of Financial Econometrics 10 (1): 54-83

[12] Engle R. F and Magdalena E. Sokalska (2012) Forecasting intraday volatility in the US equity market. Multiplicative component GARCH, Journal of Financial Econometrics 10 (1): 54-83

[13] Davidson James (2000) Econometric Theory, Basil Blackwell Publishers.

[14] Davidson R and MacKinnon J. G. (2004) Econometric Theory and Methods, Oxford.

[15] Doornik J A and D.F. Hendry ((2012) PC-Give Volume I-III, GiveWin Timberlake Consultants Limited, London.

[16] Enders W (2010) Applied Econometric Time Series, 3rd ed. John Wiley

[17] Granger C.W.J. (1980) Forecasting in Businss and Economics, Academic Press.

[18] Granger C.W.J. and P Newbold (1986) Forecasting Economic Time Series, Academic Press.

[19] Griffiths WE, Hill RC & Judge GG (1993), Learning and Practising Econometrics, J Wiley and Sons.

[20] Gabriel, V. J. and Martins, L. F. (2004), On the forecasting ability of ARFIMA models when infrequent breaks occur. The Econometrics Journal, 7: 455–475.

[21] Fawcett N, G. Kapetanios, J. Mitchell and S. Price (2013) Generalised Density Forecast Combinations, Bank of England.

[22] Hamilton JD (1994) Time Series Analysis, Princeton.

[23] Harris R and R Sollis (2003) Applied Time Series Modelling and Forecasting, John Wiley and Sons.

[24] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[25] Harvey A.C. (1989) Forecasting structural time series models and Kalman Filter, Cambridge University Press.

[26] Hendry David (1995) Dynamic Econometrics, Oxford University Press.

[27] Holly S. and M Weale ed. (2000) Econometric Methods: Practice and Applications, Cambridge University Press.

[28] Hubbard TP, T Li b , H. J. Paarsch (2012) Semiparametric estimation in models of first-price, sealed-bid auctions with affiliation, Journal of Econometrics 168 4–16

[29] Judge G G, R.C. Hill, W.E. Griffiths, H Lutkepohl and TC Lee (1988) Introduction to the Theory and Practice of Econometrics, John Wiley

[30] Koop, G. and Potter, S. (2004), Forecasting in dynamic factor models using Bayesian model averaging. The Econometrics Journal, 7: 550–565.

[31] Koop G. and S. M. Potter (2007) Estimation and Forecasting in Models with Multiple Breaks Review of Economic Studies (2007) 74 (3): 763-789

[32] Koop, G (2008) Introduction to econometrics,Chichester: Wiley.

[33] Mills TC (2003) Modelling Trends and Cycles in Economic Time Series, Palgrave

[34] Mills TC (2000) Econometric Modelling of Financial Time Series, 2nd ed. Cambridge

[35] Min CK and A Zellner (1993) Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates ,Journal of Econometrics, 56, s 1–2, March 1993, 89-118

[36] Ng, S. and Vogelsang, T. (2002), Forecasting autoregressive time series in the presence of deterministic components. The Econometrics Journal, 5: 196–224

[37] Pesaran H.M, D Pettenuzzo, and A. Timmermann (2006) Forecasting Time Series Subject to Multiple Structural Breaks Review of Economic Studies (2006) 73 (4): 1057-1084

[38] Pesaran H.M, A. Pick and A. Timmermann (2011) Variable selection, estimation and inference for multi-period forecasting problems Journal of Econometrics, 164, 1, 1 September 2011, 173-187

[39] Pyndick RS and DL Rubinfeld (1998) Econometric Models and Economic Forecasts, McGraw Hill

[40] Polito Vito and Peter Spencer (2012) UK Macroeconomic Volatility and the Welfare Costs of Inflation, Cardiff Economics Working Papers, E2011/23.

[41] Rossi B and T. Sekhposyan (2011) Understanding models' forecasting performance Journal of Econometrics, 164, 1, 1 , 158-172

[42] Singleton K. J. (2006) Empirical Dynamic Asset Pricing, Princeton.

[43] Smith P, A. Clare, J. Seaton and S. Thomas (2014) Trend following, risk parity and momentum in commodity futures, International Review of Financial Analysis, 31, 1-12.

[44] Smith P and A M Velázquez (2013) Expected returns and the business cycle: the role of supply and demand shocks, Manchester School, 81, 100-124.

[45] Stock J. H. and M. Watson (2002) Forecasting Using Principal Components from a Large Number of Predictors, Journal of the American Statistical Association, 97, 1167-1179.11

14.1 SVM (Support Vector Machine)

Support vector machines (SVMs) are a set of supervised learning methods used for classification,regression and outliers detection. It is one of the important tools in data science.

The advantages of support vector machines are:

- Effective in high dimensional spaces.
- Still effective in cases where number of dimensions is greater than the number of samples.
- Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
- Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

The disadvantages of support vector machines include:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation (see Scores and probabilities, below).

Instructions for Python programmes see:

https://scikit-learn.org/stable/modules/svm.html

References

[1] Plakandaras, V., Gupta, R., Gogas, P. and Papadimitriou, T., 2015. Forecasting the US real house price index. Economic Modelling, 45, pp.259-267.

[2] Gogas, P., Papadimitriou, T. and Takli, E., 2013. Comparison of simple sum and Divisia monetary aggregates in GDP forecasting: a support vector machines approach. Economics Bulletin, 33(2), pp.1101-1115.

[3] Plakandaras, V., Papadimitriou, T. and Gogas, P., 2015. Forecasting daily and monthly exchange rates with machine learning techniques. Journal of Forecasting, 34(7), pp.560-573.

[4] Gogas, P., Papadimitriou, T. and Chrysanthidou, E., 2015. Yield curve point triplets in recession forecasting. International Finance, 18(2), pp.207-226.

14.1.1 Extreme-Bound Analysis

Under the traditional econometric approach an investigator relies on correct sign of coefficients, significance of t-values, and high R-square in order to determine the accuracy of the model specification with no role for prior beliefs as an initial point for such specification. Only selective results that fulfil above criteria are reported in practice. EBA explicitly incorporates prior information and has systematic approach to test fragility of coefficients being reported (Leamer (1983) and Leamer and Leonard (1983) and Granger and Uhlig (1990)) as noted in the EBA algorithm in the box below.

In brief, Algorithm for the extreme bound is as proposed by Granger and Uhlig (1990) is

Let y = X■ + ■ and focus coefficient ■$_0$ = ■'■ with linear constraints C■ = c and M(C■-c) = 0. Start with a GLS estimator of the full model of ■ as b=(X'■$^{-1}$X)$^{-1}$(X'■$^{-1}$ X) then estimate of ■$_0$,

b_0=[redacted]'b with the covariance matrix of b as D=[redacted]^2 (X'[redacted]$^{-1}$X)$^{-1}$ and its decomposition A = CDC' with A$^{-\frac{1}{2}}$=A[redacted]^{1} for given M define W=A$^{\frac{1}{2}}$M'. Define two important vectors u = A$^{-\frac{1}{2}}$CD[redacted] and v =A$^{-\frac{1}{2}}$(Cb-c) and with Euclidian norm [redacted]u[redacted]= [redacted]u'u)$^{\frac{1}{2}}$ for [redacted] [0,(2/[redacted])] and cos2[redacted] = cos(u,v) [redacted] ((u'u))/([redacted]u[redacted]v[redacted]). cos2[redacted] = cos(u,v) = 0 if u = 0 or v = 0. The GLSE of [redacted] of [redacted]$_0$ under restriction M(C[redacted] - c) = 0 is [redacted]$_0$ = b_0 - u'W(W'W)[redacted]^{1}W'v . The extreme values of [redacted]$_0$ over all choices of M (full row rank) are b_0-cos^2 [redacted]u[redacted]v[redacted]<[redacted]<b_0+sin^2[redacted]u[redacted]v[redacted] or b_0-[redacted]$_L$ [redacted]u[redacted]v[redacted]<[redacted]<b_0+[redacted]$_U$ [redacted]u[redacted]v[redacted]

While the regression analysis provides points or interval estimations of coefficients for each of these explanatory variables, the EBA provides minimum and maximum effects of a variable on stock returns by lower and upper-bounds of alternations coefficients of the free variables (in this case corruption) due to inclusion or exclusion of each of the doubtful variables in the model. Ususally EBA is prorgramed in STATA or R.

- Granger, C.W. & Uhlig, H.F., 1990. Reasonable extreme-bounds analysis. Journal of Econometrics, 44(1-2), pp.159-170.
- Leamer, Edward E, (1983). Let's Take the Con Out of Econometrics, American Economic Review, American Economic Association, vol. 73(1), 31-43
- Leamer, E. and Leonard, H., 1983. Reporting the fragility of regression estimates. The Review of Economics and Statistics, pp.306-317.

15 Tutorials

Econometric Analysis

Tutorial 1: Basics
Optimisation and Matrix

Q1. Consider y as a function of x_1, x_2 and x_3 as given in the following equation:

$$y = -5x_1^2 + 10x_1 + x_1x_2 - 2x_2^2 + 4x_2 + 2x_2x_3 - 4x_3^2 \tag{1840}$$

a. Find the optimal values of , and using the first order conditions for unconstrained maximisation. Use matrix approach in your solution.

b. Determine whether the above solutions correspond to the minimum or the maximum point using positive or negative definite concepts of the Hessian determinants.

Q2. Consider coefficients of a market model given by a matrix $A = \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}$

a) What are the eigen values of this maxtrix?

b) What are associated eigen vectors?

c) Prove that eigen vectors are orthogonal, $(V_1)^{'}(V_2) = 0$

d) Prove $(V_1V_2)^{'}(V_1V_2) = (V_1V_2)(V_1V_2)^{'} = I$

Q3. Explain each of the following concepts.

a. Order of integration and unit root test.

b. Engle-Granger Representation theorem.

c. Johansen test for cointegrating vector.

d. Simultanety bias.

e. Autocorrelation.

Tutorial 2
Basic Regression Techniques

Q1. Consider a standard regression model $Y = \beta X + e$ where Y is $T \times 1$ vector of dependent variable, X is $T \times K$ matrix of explanatory variables, e is $T \times 1$ vector of independently and identically distributed normal random variable with mean equal to zero and a constant variance, that is $e \sim N(0, \sigma^2 I)$. Here β is a $K \times 1$ vector of unknown coefficients.

a. Show how best, linear and unbiased parameters $\theta = \left(\beta, \sigma^2\right)$ can be estimated using the OLS method.

b. Show briefly how the generalised least square method can be used to avoid heteroskedasticity or autocorrelation problems.

c. Prove that the ML estimators of β, σ^2 are equivalent to the OLS estimators, where:

$$\text{ML: } lnL(\theta/Y) = \ln\left\{\prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(Y-\beta X)^2}{\sigma^2}\right]\right\} \tag{1841}$$

d. Show that the GMM estimators of β, σ^2 are equivalent to both the OLS and ML estimators.

e. Why are GMM and ML estimators more popular in advanced studies than the OLS estimators? Comment with some examples.

Q2. Consider the maximum likelihood function given below:

$$\text{ML: } lnL(\theta/Y) = \ln\left\{\prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(Y-\beta X)^2}{\sigma^2}\right]\right\} \tag{1842}$$

a. How are parameters α, β, σ^2 estimated in this model?

b. For a linear function , prove that ML estimators of α, β, σ^2 are equivalent to the OLS estimators.

c. Discuss differences between the likelihood ratio test, Lagrange multiplier test and the Wald test. Use diagrams and equations to illustrate your answer.

d. Illustrate how the maximum likelihood method can be applied in estimating parameters in a ARMA(1,1) or ARCH(p,q) or GARCH(p,q) model.

Tutorial 2. 1
Simultaneous Equation

Q1. Consider a market model for a particular product.

Demand: $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t}$ (1)

Supply: $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t}$ (2)

Here Q_t^d is quantity demanded and Q_t^s is quantity supplied, P_t is the price of commodity,P_{t-1} is price lagged by one period, I_t is income of an individual, $u_{1,t}$ and $u_{2,t}$ are independently and identically distributed (iid) error terms with a zero mean and a constant variance.Q_t and P_t are endogenous variables and $P_{t-1}$and I_t are exogenous variables α_0,α_1,α_2,and β_0,β_1,$\beta_2$are six parameters defining the system.

1. How can simultaneity bias occur if one tries to apply OLS to each of the above equations.

   Use rank and order conditions to judge whether each of these two equations are over-, under- or exactly identified.

2. Write down the reduced form for this system.

3. How would you estimate the coefficients of the reduced form equations? Write down the estimator.

4. If equations are identified explain how you may retrieve the structural parameters α_0,α_1,α_2,and β_0,β_1,β_2, and from the coefficients of the reduced form equations.

Q2. Consider Keynes-Hicks Macroeconomic Model: Goods Market
Consumption function

$$C_t = \beta_0 + \beta_1 (Y_t - T_t) + \beta_2 X_t + \varepsilon_{1,t} \tag{1843}$$

Taxes:

$$T_t = t_0 + t_1 Y_t + t_2 M_t + t_3 G_t + \varepsilon_{2,t} \tag{1844}$$

Imports:

$$M_{\mathbf{t}} = \; m_0 + m_1 Y_t + m_2 M_t + m_3 G_t + \varepsilon_{3,t} \tag{1845}$$

Investment

$$I_t = \mu_0 - \mu_1 R_t + \phi \Delta Y_{t-1} + \varepsilon_{4,t} \tag{1846}$$

Keynes-Hicks Macroeconomic Model: Money Market

$$\left(\frac{\overline{MM}}{P}\right)_t = \; b_0 + b_1 Y_t - b_2 R_t + \varepsilon_{5,t} \tag{1847}$$

Macro balance

$$Y_t = \; C_t + I_t + G_t + X_t - M_t = C_t + T_t + S_t \tag{1848}$$

Money Market Equilibrium

$$R_t = \frac{b_0}{b_2} - \frac{1}{b_2}\left(\frac{\overline{MM}}{P}\right)_t + \frac{b_1}{b_2} Y_t \tag{1849}$$

Here Y_t, R_t, C_t, I_t, T_t, M_t are endogenous variables (income, interest rate, consumption, investment, tax revenue and imports);
government spending, exports, real money balances-liquidity are exogenous policy variables, G_t, X_t, $\left(\frac{\overline{MM}}{P}\right)_t$; behavioral parameters are β_0, β_1 $T_0, \mu_0, m_0, t_1, m_1, b_0, b_1, b_2$.
a. Construct a table of structural coefficients in the tabular form.
b. Check whether each of the equation is identified or not using rank and order conditions.
c. Write the reduced form of the model.
d. Estimation the model using data in macro08.csv.

e. Do historical simulations using this model.

f. Project the values of exogenous variables for next 10 quarters.

g. Forecast the economy for next ten quarters using estimated models and projected values of exogenous variables.

Tutorial 3
VAR and cointegration analysis

Q1. Consider a structural VAR model between y_t and z_t as following:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1850}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1851}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$.

a. Derive the reduced form of this VAR model and suggest ways to estimate the parameters in it.

b. How would one determine stability of a VAR system like this? Provide analytical solutions using the roots of the quadratic function.

c. How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified?

d. Write impulse response functions for these two equations and indicate how can one perform an impulse response analysis with them?

e. What is the meaning of variance decomposition in a VAR model like this?

Q2. Consider a vector error correction model (VECM) of the form

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1852}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_3 + ... + \Pi_i \quad$ for $i = 1..k \quad \Gamma_k$ gives the long run solution and $\varepsilon_t \sim N(0, \sigma_y^2)$.

a. What is the meaning of cointegration and why should there exist at least one cointegrating vector in this equation?

b. Discuss how likelihood ratio tests are employed to determine the optimal lag.

c. Explain how the canonical correlations provide eigenvalues and eigen vectors that are useful in determining the rank of the cointegrating vector.

d. Discuss a procedure for trace and max-eigenvalue tests for cointegration.

Tutorial 4
Dynamic Analysis and the Maximum Likelihood

Q1. Consider the Full Information Maximum Likelihood (FIML) method to estimate a standard macroeconomic model of the form

$$C = a_0 + a_1(Y - T) + a_2 Y_{t-1} + u_1 \tag{1853}$$

$$I = b_0 + b_1 Y + b_2 K_{t-1} + b_3 r + b_4 E + u_2 \tag{1854}$$

$$M = c_0 + c_1 Y + c_2 P_{t-1} + u_3 \tag{1855}$$

$$Y = C + I + G + E - M \tag{1856}$$

where endogenous variables Y, C, I, M are income, consumption, investment and imports respectively; exogenous variables $G, E, T, K_{t-1}, P_{t-1}$ and r denote government spending, exports, tax revenue, lagged capital, lagged price level and interest rate respectively; $u_1, u_2$and u_3 are error terms in consumption, investment and imports respectively. Parameters a_0, a_1, a_2, b_0, b_1, b_2, b_3, b_4, c_0, c_1 and c_2 provide behavioral relations in the model.

a. Derive the reduced form coefficients of this system and write the reduced form equations in terms redefined endogenous variables $y_1 = C$; $y_2 = I$; $y_3 = M$ and exogenous variables $z_1 = G; z_2 = E; z_3 = K_{t-1}; z_4 = T; z_5 = Y_{t-1}; z_6 = r; z_7 = P_{t-1}$.

b. Write expressions for the joint distribution of errors, $u_{11}u_{12}u_{13}......u_{1n}; u_{21}u_{22}u_{23}......u_{2n}; u_{31}u_{32}u_{33}......u_{3n}$, in this system

c. Derive the Jacobian matrix of the first order conditions to estimate the parameters of the reduced system

$$u_1 = y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5 \tag{1857}$$

$$u_2 = y_2 - \beta_0 - \beta_1(y_1) + \beta_1(y_3) - \beta_1(z_1) - \beta_2 z_3 - \beta_3 z_6 - \beta_4 z_2 \tag{1858}$$

$$u_3 = y_3 - \gamma_0 - \gamma_1(y_1) - \gamma_1(y_2) - \gamma_1(z_1) - \gamma_1(z_2) - \gamma_2 z_7 \tag{1859}$$

d. Derive the log likelihood functions for y variables using the transformation functions as:

$$P(y_{1i}.y_{2i}.y_{3i}) = P(u_{1i}.u_{2i}.u_{3i}) . \left| \frac{\partial(u_{1i}.u_{2i}.u_{3i})}{\partial(y_{1i}.y_{2i}.y_{3i})} \right| \tag{1860}$$

e. Show first order conditions for maximisation of the log likelihood function in terms of the reduced form parameters α_0, $\alpha_1, \alpha_2, \beta_0$, $\beta_1, \beta_2, \beta_3, \beta_4, \gamma_0$, γ_1 and γ_2 .

f. Explain procedure on how all FIML parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2$ and $\sigma^2_{u_1}, \sigma^2_{u_2}$and $\sigma^2_{u_3}$ could be estimated and how the log FIML could be evaluated.

g. Why is FIML the most efficient estimation technique for a system like this?

Q2. Consider a dynamic model of Y_t on Y_{t-1}, X_t and X_{t-1} in the **LSE** tradition as given below.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \tag{1861}$$

a. What is the error correction form of this model?

b. Briefly analyse the implications of following restrictions in the comparative static or dynamic properties of the model.

1) $\alpha_1 = 0; \beta_1 = 0$
2) $\beta_1 = 0$
3) $\alpha_1 = 0; \beta_0 = 0$
4) $\alpha_1 = 0$
5) $\alpha_1 = 1$; $\beta_0 + \beta_1 = 0$
6) $\beta_0 = 0$

c. An urn contains N balls and N1 of them are red. The likelihood function of being red in N_1 draws is given by the discrete Bernauli probability distribution function as

$$p(N) = p^{N_1}(1-p)^{N-N_1} \tag{1862}$$

Using the maximum likelihood estimatore i) prove that $p = \frac{N_1}{N}$ and ii) it maximises the likelihood function (second order conition).

Q3. Consider a Lagrange function for a restricted least square model.

$$\begin{aligned} L &= e'e + 2\lambda\left(r' - \beta' R'\right) = (Y - \beta X)'(Y - \beta X) + 2\lambda\left(r' - \beta' R'\right) \\ &= Y'Y - 2\beta X'Y + (\beta X)'(\beta X) + 2\lambda\left(r' - \beta' R'\right) \end{aligned} \tag{1863}$$

a. Using the first order conditions derive the estimator of parameter veror (b) given by the restricted least square. What is the value of λ?

b. Prove that the variance of the restricted least square is smaller than the variance of the unrestricted least square.

Tutorial 5
Time Series, ARMA, ARIMA

Q1. Consider a monthly time series $\{y_t\}$

a. Show how the traditional moving average based methods could be applied to decompose its trend, seasonal, cyclical and irregular components.

b. Consider a random walk model $y_t = y_{t-1} + \varepsilon_t$ with initial conditions $y_1 = y_0$ for t =1.What are the mean, variance and the time path of y_t in terms of current and past series of errors ε_t ? What is its conditional forecast for period j made at time t? What is the error of forecast and its variance? How are the mean and variances affected if this random walk includes a drift term a_0 as in $y_t = y_{t-1} + a_0 + \varepsilon_t$.

c. Consider signal extraction problem for series y_t including permanent and transitory shocks components as ε_t and η_t

$y_t = \varepsilon_t + \eta_t$ and $\varepsilon_t^* = a + by_t$

where $E(\varepsilon_t) = 0$; $E(\eta_t) = 0$;$E(\varepsilon_t \eta_t) = 0$; $E(\varepsilon_t^2) = \sigma^2$; $E(\eta_t^2) = \sigma^2$.

What is its minimum square error (MSE)? How is the partitioning parameters b optimally estimated?

d. What are the prominent reasons for a failure of forecast? Illustrate Ganger and Newbold (1986) technique for combining optimal forecasts as in $f_{ct} = (1-\lambda) f_{1t} + \lambda f_{2t}$.

Q2. What is the main principle of forecasting and what are the reasons for failure of model based forecasts? Derive the forecast errors and variance of forecast for the following forecasting models .

a. Random walk with a drift: $\left[y_1 = y_0 + a_0 + \epsilon_1,\ e_{T+1} \sim N(0,1)\right]$.

b. Period h ahead forecast of AR(1): $\left[y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h},\ e_{T+h} \sim N(0,1)\right]$.

c. One period ahead forecast in MA(1): $\left[y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T,\ e_{T+1} \sim N(0,1)\right]$.

d. Two period ahead forecast in ARMA(1,1): $\left[y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1}, e_{T+2} \sim N(0,1)\right]$.

Tutorial 6
Cross section analysis: Probability Models

Q1. Discuss the maximum likelihood functions and Newton-Ralphson or BHHH algorithms for estimation of parameters and the testing procedure for the following cross section models:

a. Logit $\left[P_i = \frac{1}{1+e^{-Z_i}} \text{ with } Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i\right]$.

b. Count data $\left[P\left(Y=y\right) = \frac{e^{-\lambda}\lambda^y}{y!}\right]$.

c. Multinomial Choice model: $\left[\frac{P_{i,2}}{P_{i,1}} = \frac{\frac{\exp\left(X'_{i,2}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}}{\frac{\exp\left(X'_{i,1}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}} = \frac{\exp\left(X'_{i,2}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)}\right]$.

d. Ordered probit model: $\left[prob\left(y=J|x\right) = 1 - \Phi\left(\mu_{J-1} - x'\beta\right)\right]$.

e. Heckman's correction for selectivity bias in which

$Y_{1,i} = X_{1i}\beta + \varepsilon_{1,i}$ and $Y_{2,i} = X_{2,i}\beta + \varepsilon_{2,i}$ and if the event occurs $Y_{1,i} > Y_{2,i}$.

f. Two limit Tobit for a certain regression $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ with

$$Y_i = \left(\begin{array}{l} Y_i^* \quad if \quad L_1 < \ Y_i^* < L_2 \text{ if the event occurs} \\ L_1 \quad \text{if } Y_i^* < L_1 \\ L_2 \quad \text{if } Y_i^* \geq L_2 \end{array}\right).$$

Tutorial 7
Panel Data Models

Q1. Consider a panel data set with time $t = 1, .., .., .T$ and individuals $i = 1, \ldots .N$. Discuss the specification, estimation and testing procedure for the following versions of panel data models

a. Pooling cross section and time series in SUR model.

b. Fixed effect model for $\left[y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t} \qquad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)\right]$.

c. Random effect model for $\left[y_{i,t} = \alpha_i + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \qquad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)\right]$.

d. Dynamic panel data model $\left[y_{i,t} = \alpha_i + \theta_i y_{i,t-1} + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)\right]$.

e. Im-Pesaran-Shin and KPSS panel unit root tests.

f. Pedroni's panel cointegration test.

Tutorial 8
Duration Analysis

Q1. Derive duration density, hazard rate, survival function and duration dependence for the following duration or hazard functions and explain the general procedure for estimation of model parameters.

a. Exponential distribution. $\left[F(t) = 1 - \exp(-\lambda.t)\right]$.

b. Weibull $\left[f(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha})\right]$.

c. Log normal distribution $\left[f(t) = \frac{1}{\sigma.t}\phi\left(\frac{\log T - m}{\sigma}\right)\right]$.

d. Log logistic $\left[f(t) = \alpha\gamma t^{\alpha-1}\left(1 + \gamma t^{\alpha}\right)^{-2}\right]$.

e. Gamma distribution $\left[f(t) = \frac{\left[a^v t^{v-1}\exp(-at)\right]}{\Gamma(v)} \quad \text{where } \Gamma(v) = \int_0^{\infty}\exp(-t)\, t^{v-1}\partial s.\right]$

Tutorial 9
Bayesian Modelling

Q1. Write short notes on any five of the following issues relating to the Bayesian modelling and analysis.

a. Difference between classical and Bayesian assumptions on parameters and errors in a linear regression.

b. Bayesian rule where $p(A, B)$ denotes the joint probability of occurring events A and B together, $p(B)$ is the marginal probability of B without any respect to occurrence of A. Similarly $p(A)$ is the marginal probability of A without any respect to occurrence of B.

c. Bayesian prior and posterior density functions on unknown parameters β and τ for the likelihood function such as: $\log l(\beta, \tau) = \frac{1}{2}[n \log \tau - \tau (\beta - b) X'X (\beta - b) - \tau e'e]$

d. Estimates of the mean and variance of β and τ in the above Bayesian linear regression model.

e. Bayesian panel data model of the form $y_{i,t} = x_{i,t}\beta + \alpha_i + e_{i,t}$ with priors $e_{i,t}/x_{i,t}, \beta, \alpha_i, \tau \sim N(0, \tau)$.

f. Estimation procedure in a Bayesian VAR model of the form: $\underset{T\times 1}{Y_j} = \underset{T\times K}{X_j} \underset{K\times 1}{\beta_j} + \underset{T\times 1}{u_j}$ where $u_j \sim N\left(0, \sigma_{i,j}^2 I_T\right)$.

g. Specification and estimation procedure of the Bayesian stochastic volatility models (Bayesian ARCH/GARCH).

h. MCMC algorithm.

Tutorial 10
GMM

Q1. Consider consumer optimisation problem in a capital asset pricing model popularised by Hansen and Singleton (1982) stated as:

$$max\ E_0 \left[\sum_{t=0}^{\infty} \delta^t U(C_t) \right] \tag{1864}$$

subject to

$$C_t + P_t Q_t \leq R_t Q_{t-1} + W_t \tag{1865}$$

a. Formulate the generic Lagrangian function for constrained optimisation to solve this problem. Then modify the utility function as $U(C_t) = \frac{C_t^{\gamma}}{\gamma}$ with $\gamma < 1$; $\alpha = \gamma - 1$.

b. Derive the Euler equations based on the first order conditions that maximise the objective function of the model.

c. Discuss how the GMM could be applied to estimate the parameters δ and α . Why the application of the maximum likelihood is computationally cumbersome here and generates biased results?

d. Discuss properties of Arellano-Blundel-Bond GMM estimators for a dynamic panel data model to estimate such CAPM model across countries.

e. For applied studies discuss the GMM estimation procedure in Eviews/RATS/STATA/PcGive.

16 Assignment

Write an essay in 2000 words in any one of the following issues; this essay accounts for 20 percent of the module marks. For each topic first present the theoretical model and derivations. Then state major hypotheses and economic logic behind it. Select variables from the dataset; examine summary statistics of those variables. Do estimations and interpret their significance and test their validities and suggest limitations and alternative structures of the model. Compare efficiency of OLS, maximum likelihood, the GMM or non-parametric estimators for the issue at hand. Explain contribution of analysis in coming decision and relate to more advanced methods developed in recent years such as non-linear, non-parametric or dynamic factor models.

1. Do cross section analysis using the data in the annual population survey or the longitudinal dataset contained in the Understanding Society (the BHPS or the PIDS or German Social and Economic Survey or Luxumberg poverty study or similar cross section dataset of your choice).
2. Develop a Bayesian macro model for policy coordination and estimate model parameters using the MCMC and Kalman Filter algorithms. Use MATLAB/dynare.
3. Develop classical and Bayesian VAR/cointegration model for a country of your choice and apply it for policy analysis such as for assessing the macroeconomic impacts Brexit process in the UK or the EU economies.
4. Analyse how volatility of stock prices are linked to the economic performance of comporations and macroeconomic indicators using ARCH/GARCH or multivariate ARCH/GARCH models on quarterly time series data.
5. Apply static, dynamic or threshold panel data models to assess impacts of economic growth on poverty and inequality. Do open countries grow faster and achieve more equality in income distribution?
6. Is there an optimal size of public debt? Analyse this issue using a fiscal and monetary policy simulation and forecasting model for a country of your choice. Analyse impacts and implications of quantitative easing.
7. Assess impacts of tax or trade policies using a dynamic general equilibrium model of an economy. Use GAMS/MPSGE for computing such model and excel based routines in processing and presenting results in a paper.
8. Apply duration model to assess conditional hazard rates of some important economic phenomena such as end of unemployment spells, failure of a company, change in price of products, or probability of occurrence of certain medical problem such as stroke among a given population.
9. Has the price of energy been a constraint in economic growth? Use non-parametric techniques for analysis such as Support Vector Machine (SVM), LASSO or ridge regressions or Exteme Bound Analysis (EBA).
10. Quantify the impact of efficiency in the financial or logistic sectors in the economic growth rate and consumer welfare in an economy.
11. Use cross section surveys in order to study consumer or employee satisfaction in a given market or in institutions.

16.0.2 Instructions for the essay

1. Essay should contain motivation, brief review of relevant literature, analytical section, estimation or computation and conclusions, recommendations and references. You must report of validity of results using test statistics or sensitivity analysis.

2. Write in your own words referring to economic theories and evidences and literature. Both quantitative and qualitative methods can be applied according to your interest. Be critical, analytical and very precise. Submit the electronic copy of the essay through Turnitin and a hard copy to the research office by the deadline.

3. CConsult essays on economic theory and other reading materials. Construct additional reading list that might be helpful in preparation for the above assignment. Ask and sear for related articles. They can be found from JSTOR and Econlit..

4. Some examples of data and relevant software information is given in the Ebridge site. More detailed derivations could be kept in the appendix.

5. Provide the topic by the second week, specification of the model by the fourth week, estimation and analysis by the sixt week and draft of the essay by the 8th week and then submit the essay to the research office on the day mentioned in the module handbook. Materials from the lectures and tutorial should be applied and referenced as usual with other articles, tests or reports to substantiate the findings.

Assessment criteria: This is expected to be a professional piece of work and must contain a model and analysis. Be original, critical, systematic, concise, consistent, organised in presentation of your arguments. Essay should follow a style of journal article. The elements of marks will broadly be based on the originality of the question (20%) analytical structure (30%); estimation and computation (20%), explanation of results (20%), overall presentation (10%). Students are allowed to ask any question on the chosen topic in any teaching sessions.

The essay should be typed in the double space, checked for spelling, grammar and pagination. Put word counts in the top right corner of the front page. You must read the Business School policies regarding academic honesty and consequences of plagiarism as mentioned in the Business School Skills Handbook and declare academic honesty by filling in cover sheet of submission.

16.1 Econometric and Statistical Software

- Excel
- OX-GiveWin/PcGive/STAMP
- Eviews
- Shazam
- microfit
- JMulti
- RATS
- NLOGIT/ LIMDEP
- GAUSS
- STATA/SPSS
- R http://www.ats.ucla.edu/stat/r/ ;
- **http://www.feweb.vu.nl/econometriclinks/;** https://www.aeaweb.org/rfe/

1. Excel Spreadsheets are very user friendly and could be used for algebraic calculations and statistical analyses for many kinds of economic models. First prepare an analytical solution by hand then use Excel formula to compute. Excel has constrained optimiser routine at tool/goal seek and solver commend. It also contains matrix routines to get determinants of matrices and to multiply and invert them using multiple cell options (see Koop (2007)).

2. OX-GiveWin/PcGive/STAMP (www.oxmetrics.net) is a very good econometric software for analysing time series and cross section data. This software is available in all labs in the network of the university by sequence of clicks Start/applications/economics/givewin. Following steps are required to access this software.

a. save the data in a standard excel file. Better to save in *.csv format .

b. start give win at start/applications/economics/givewin and pcgive (click them separately)

c. open the data file using file/open datafile command.

d. choose PcGive module for econometric analysis.

e. select the package such as descriptive statistics, econometric modelling or panel data models.

d. choose dependent and independent variables as asked by the menu. Choose options for output.

e. do the estimation and analyse the results, generate graphs of actual and predicted series.

A Batch file can be written in OX for more complicated calculations using a text editor such as pfe32.exe. Such file contains instructions for computer to compute several tasks in a given sequence.

Doornik J A and D.F. Hendry ((2003) PC-Give Volume I-III, GiveWin Timberlake Consultants Limited, London

R : https://www.coursera.org/course/rprog

16.1.1 Mathematical software

4. GAMS is good particularly in solving linear and non-linear problems. It has widely been used to solve general equilibrium models with many linear or non-linear equations on continuous or discrete variables. It comes with a number of solvers that are useful for numerical analysis. For economic modelling it can solve very large scale models using detailed structure of consumption, production and trade arrangements on unilateral, bilateral or multilateral basis in the global economy where the optimal choices of consumers and producers are constrained by resources and production technology or arrangements for trade.

It is a user friendly software. Any GAMS programme involves

- declaration of set, parameters, variables, equations,
- initialisation of variables
- setting their lower or upper bounds
- solving the model using Newton or other methods for linear or non-linear optimisation
- and reporting the results in tables or graphs (e.g. ISLM.gms).

Full version of GAMS/MPSGE program is good for large scale standard general equilibrium models. GAMS programme can be downloaded from demo version of GAMS free from www.gams.com/download).

The check whether the results are consistent with the economic theory underlying the model such as ISLM-ASAD analysis for evaluating the impacts of expansionary fiscal and monetary policies. Use knowledge of growth theory to explain results of the Solow growth model from Solow.gms.

Consult GAMS and GAMS/MPSGE User Manuals, GAMS Development Corporation, 1217 Potomac Street, Washington D.C or www.gams.com or www.mpsge.org for GAMS/MPSGE.

For other relevant software visit: http://www.feweb.vu.nl/econometriclinks/ or https://www.aeaweb.org/rfe/;

16.1.2 MATLAB

MATLAB is widely used for solving models. It has script and function files used in computations.

Both have *.m extensions. Its syntax are case sensivite. Solving a system of linear equations and handling matrices

Example 1

Write a programme file matrix.m like the following and try run.

```
% now solve a linear equation
% 5x1 + 2x2 =20
% 3x2 + 4x2 =15
k =[5 2;3 4];
n = [20 15];
kk = inv(k)
x = kk*n'
```

One more example of system of equation and factorisation of matrices

```
A=[1 2 3; 3 3 4; 2 3 3]
b=[1; 2; 3]
%solve AX=b
```

```
X = inv(A)*b
%eigen value and eigenvectors of A
[V,D]=eig(A)
%LU decomposition of A
[L,U]=lu(A)
%orthogonal matrix of A
[Q,R]=qr(A)
%Cholesky decomposition (matrix must be positive definite)
%R = chol(A)
%Singular value decomposition
[U,D,V]=svd(A)
```

Contents.m for list of files in MATLAB demo. MATLAB demo available in http://www.youtube.com/.
Run "kalman_run.m" in MATLAB with function "kalmanf.m" from the eBridge.

Relevant web pages

http://www.khanacademy.org/ ; http://www.feweb.vu.nl/econometriclinks/
http://www.econometricsociety.org/; http://www.aeaweb.org/aer/index.php; http://www.res.org.uk/economic/ejbro
http://www.imf.org/external/pubs/ft/weo/2010/01/weodata/index.aspx;
http://www.ifs.org.uk/publications/789
http://www.esds.ac.uk/international/; http://www.bankofengland.co.uk/;
http://www.hm-treasury.gov.uk/
http://www.eea-esem.com/EEA/2010/Prog/ - look at fiscal policy sessions.

17 Articles and texts

References

[1] Andersen TG and BE Sorensen (1996) GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, Journal of Business and Economic Statistics, 44:3:Jul: 328-352.

[2] Angrist, J.D. and Pischke, J.S., 2015. Mastering Econometrics, Princeton university press

[3] Angrist, J.D. and Pischke, J.S., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton university press

[4] Gallant R A and G Tauchen (1996) Which Moments to Match? Econometric Theory, 12:4:Oct: 657-681.

[5] Greene W. (2008) Econometric Analysis, Sixth Edition,Pearson Education, New York.

[6] Hall A (2003) Generalised Methods of Moments in Baltagi ed. Theoretical Econometrics, Blackwell.

[7] Hall A (1993) Some aspects of generalised method of moemtns estimation, Maddala, Rao and Vinod eds., Handbook of Statistics, 11,pp.393-417.

[8] Hansen L.P. (1982) Large sample properties of generalized method of moment estimators, Econometrica, 50:4:1029-1054.

[9] Hansen L.P. and K J Singleton(1982) Generalized Instrumental Variables Estimation of Non-linear Rational Expectations Models, Econometrica, 50:5:1269-1286.

[10] Hansen L.P. J. Heaton and A Yaron (1996) Finite Sample Properties of Some Alternative GMM Estimators, Journal of Business and Economic Statistics, 14:3:Jul: 262-280.

[11] Hamilton JD (1994) Time Series Analysis, Princeton.

[12] Hendrry David (1995) Dynamic Econometrics, Oxford University Press.

[13] Imbens, G. (2000). 'The Role of the Propensity Score in Estimating Dose-Response Functions'. Biometrika, 87(3): 706–10.

[14] Imbens, G. (2000) 'The Role of the Propensity Score in Estimating Dose-Response Functions'. Biometrika 87(3): 706-10

[15] Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment effects. Econometrica, 62(2), 467–475

[16] Newey W K (1988) Asymptotic Equivalnce of Closet Moments and GMM Estimators,Econometric Theory, 4:2:Aug:336-340.

[17] Newey W K and R J Smith (2004) Higher order properties of GMM and Eneralised Likelihood Estimators, Econometrica, 72:1:Jan:219-255.

[18] Ruud P A (2000) An Introduction to Classical Econometric Theory, Oxford.

[19] Smith R J (1997) Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation, Economic Journal 107:March, 503-519

[20] Tauchen G (1986) Statistical properties of generalized methods of moments estimators of structural parameters obtained from financial marekt data, Journal of Business and Economic Statistics, 4:4:Oct: 397-416.

[21] Verbeek M. (2004) A Guide to Modern Econometrics, 2nd edition, Wiley.

[22] Wright J H (2003) Detecting Lack of Identification in GMM, Econometric Theory, 19:2:Apr: 322-330.

[23] Mills T.C. (2003) Modelling Trends and Cycles in Economic Time Series, Palgrave.

[24] Mills T.C. (2000) The Econometric Modelling of Financial Time Series, Cambridge University Press.

[25] Newey W K (1988) Asymptotic Equivalnce of Closet Moments and GMM Estimators,Econometric Theory, 4:2:Aug:336-340.

[26] Newey W K and R J Smith (2004) Higher order properties of GMM and Eneralised Likelihood Estimators, Econometrica, 72:1:Jan:219-255.

[27] Ruud P A (2000) An Introduction to Classical Econometric Theory, Oxford.

[28] Singleton K. J. (2006) Empirical Dynamic Asset Pricing, Princeton.

[29] Smith R J (1997) Alternative Semi-Parametric Likelihood Approaches to Generalised Method of Moments Estimation, Economic Journal 107: March, 503-519

[30] Tauchen G (1986) Statistical properties of generalized methods of moments estimators of structural parameters obtained from financial marekt data, Journal of Business and Economic Statistics, 4:4:Oct: 397-416.

[31] Verbeek M. (2012) A Guide to Modern Econometrics, 3rd edition, Wiley.

[32] Wright J H (2003) Detecting Lack of Identification in GMM, Econometric Theory, 19:2:Apr: 322-330.

17.1 Professional articles in econometrics

References

[1] Abadie, A. and Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. econometrica, 74(1), pp.235-267

[2] Athey, S. and Imbens, G.W., 2019. Machine learning methods that economists should know about. Annual Review of Economics, 11, pp.685-725.

[3] Imbens, G.W. and Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. Journal of economic literature, 47(1), pp.5-86.

[4] Imbens, G.W., 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. Review of Economics and statistics, 86(1), pp.4-29.

[5] Imbens, G.W. and Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. Journal of econometrics, 142(2), pp.615-635.

[6] Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C. and Cesarini, D., 2018. Redefine statistical significance. Nature human behaviour, 2(1), pp.6-10.

[7] Anderson T.W. (1996) R.A. Fisher and Multivariate Analysis, Statistical Science, 11:1:20-34.

[8] Bartlett M.S. (1947) Multivariate Analysis, Supplement to the Journal of the Royal Statistical Society, 9:2:176-197.

[9] Beyer A, REA Farmer, J Henry and M. Marcellino (2005) Factor Analysis in a New Keynesian Model, European Central Bank, Working Paper Series, 510, August

[10] Beveridge S. and C. R. Nelson (1981)A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', Jounral of Monetary Economics, 7:151-174

[11] Baltagi Badi H. (1992) Sampling Distributions and Efficiency Comparisons of OLS and GLS in the Presence of Both Serial Correlation and Heteroskedasticity Econometric Theory, 8, 2 (Jun., 1992), pp. 304-305

[12] Blomquist N. Soren (1980) A Note on the Use of the Coefficient of Determination The Scandinavian Journal of Economics, 82, 3 (1980), pp. 409-412

[13] Blundell Richard and Ian Preston (1998) Consumption Inequality and Income Uncertainty The Quarterly Journal of Economics, 113, 2 (May, 1998), pp. 603-640

[14] Blundell Richard and Costas Meghir (1986) Selection Criteria for a Microeconometric Model of Labour Supply Journal of Applied Econometrics, 1, 1 (Jan., 1986), pp. 55-80

[15] Box G and G Jenkins (1976) Time Series Analysis, Forecasting and Conctrol, Holden Day.

[16] Bauwens L , M. Lubrano and J. F. Richard (1999) Bayesian Inference in Dynamic Econometric Models, Oxford University Press.

[17] Camba-Mendez G, G. Kapetanios, R. J. Smith, Martin R. Weale (2003) Tests of Rank in Reduced Rank Regression Models, Journal of Business & Economic Statistics, 21, 1, 145-155

[18] Camba-Mendez G, G. Kapetanios, R. J. Smith, Martin R. Weale (2001) An automatic leading indicator of economic activity: forecasting GDP growth for European countries, The Econometrics Journal, 4, 1, S56-S90

[19] Card, David. "Immigration and inequality." American Economic Review 99, no. 2 (2009): 1-21

[20] Chesher A (2010) Instrumental Variable Models for Discrete Outcomes, Econometrica, 78, 2, 575–601

[21] Chesher, A. D. (2003): "Identification in Nonseparable Models," Econometrica, 71, 1405–1441

[22] Chesher A (1984) Improving the efficiency of Probit estimators, Review of Economic Studies,66:3:523-527.

[23] Chesher, A. and Smith, R.J., 1997. Likelihood ratio specification tests. Econometrica: Journal of the Econometric Society, pp.627-646.

[24] Cochrane D and G. H. Orcutt (1949) Application of Least Squares Regression to Relationships Containing Auto- Correlated Error Terms Journal of the American Statistical Association, 44, 245 (Mar., 1949), pp. 32- 61

[25] Davidson Russell, James G. Mackinnon, Russel Davidson (1985) Heteroskedasticity-Robust Tests in Regressions Directions: Annales de l'inséé, 59/60, Économétrie non linéaire asymptotique (Jul. - Dec., 1985), pp. 183-218

[26] Dickey D.A. and W.A. Fuller (1979) Distribution of the Estimator for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74:366:427-437,June.

[27] Durbin J. and G. S. Watson (1950) Testing for Serial Correlation in Least Squares Regression: I, Biometrika, 37, 3/4 (Dec., 1950), pp. 409-428

[28] Durbin J. (1987) Statistics and Statistical Science Journal of the Royal Statistical Society. Series A (General), 150, 3 (1987), pp. 177-191

[29] Durbin J. and M. G. Kendall (1951) The Geometry of Estimation Biometrika, 38, 1/2 (Jun., 1951), pp. 150-158

[30] Durbin J. and G. S. Watson (1951) Testing for Serial Correlation in Least Squares Regression. II Author(s): Source: Biometrika, 38, 1/2 (Jun., 1951), pp. 159-177

[31] Durbin J. (1970) Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables Econometrica, 38, 3 (May, 1970), pp. 410-421

[32] Durbin J. and G. S. Watson (1971) Testing for Serial Correlation in Least Squares Regression. III Biometrika, 58, 1 (Apr., 1971), pp. 1-19

[33] Engle R E and C.W.J. Granger (1987) Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica, 55, 2, 251-276.

[34] Farrar Donald E. and Robert R. Glauber (1967) Multicollinearity in Regression Analysis: The Problem Revisited The Review of Economics and Statistics, 49, 1 (Feb., 1967), pp. 92-107

[35] Fisher R. A. (1923) Statistical Tests of Agreement between Observation and Hypothesis, Economica, 8 , 139-147

[36] Fisher Jonas D.M. and Ryan Peters (2010) Using stock returns to identify Government spending shocks, The Economic Journal, 120 (May), 414–436.

[37] Garratt A., K. Lee, M.H. Pesaran and Y. Shin (2003) A Structural Cointegration VAR Approach to Macroeconometric Modelling, Economic Journal,113:487:412-455.

[38] Glejser H. (1969) A New Test for Heteroskedasticity Journal of the American Statistical Association, 64, 325 (Mar., 1969), pp. 316- 323

[39] Godfrey L. G. and M. R. Wickens (1981) Testing Linear and Log-Linear Regressions for Functional Form The Review of Economic Studies, 48, 3 (Jul., 1981), pp. 487-496

[40] Griffiths W. E. and R. C. Hill and G. G. Judge (1993) Learning and Practicing Econometrics, John Willey

[41] Grubb David and Lonnie Magee (1988) A Variance Comparison of OLS and Feasible GLS Estimators Econometric Theory, 4, 2 (Aug., 1988), pp. 329-335

[42] Hall Alastair R. (2000) Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test Econometrica, 68, 6 (Nov., 2000), pp. 1517-1527

[43] Hall Alastair (1987) The Information Matrix Test for the Linear Model,The Review of Economic Studies, 54, 2 (Apr., 1987), pp. 257-263

[44] Hansen P. L. (1982) Large sample properties of generalised methods of moments estimators, Econometrica, 50:4:1029-1054.

[45] Harvey Andrew (1997) Trends, Cycles and Autoregressions, The Economic Journal, 107, 440 (Jan., 1997), pp. 192-201

[46] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[47] Hansen L.P. (1982) Large sample properties of generalized method of moment estimators, Econometrica, 50:4:1029-1054.

[48] Hamilton James D. (1994) Time Series Analysis, Princeton.

[49] Hausman J.A., (1978), Specification Tests in Econometrics, Econometrica, 46, 6, pp.1251-1271.

[50] Hausman Jerry A. (1975) An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models Econometrica, 43, 4 (Jul., 1975), pp. 727-738

[51] Hausman Daniel M. (1989) Economic Methodology in a Nutshell The Journal of Economic Perspectives, 3, 2 (Spring, 1989), pp. 115-127

[52] Heckman James J. (1978) Dummy Endogenous Variables in a Simultaneous Equation System Econometrica, 46, 4 (Jul., , pp. 931-959

[53] Heckman J. J., (1979) Sample Selection Bias as a Specification Error, Econometrica, 47, 1, pp153-161.

[54] Hendry D.F. (1997) The Econometrics of Macroeconomic Forecasting , *Economic Journal*, 107, 444., 1330-1357

[55] Hendry D (1995) Dynamic Econometric Theory, Oxford.

[56] Hendry David F. (1974) Stochastic Specification in an Aggregate Demand Model of the United Kingdom Econometrica, 42, 3 (May, 1974), pp. 559-578

[57] Imbens G. W. and T Lancaster (1994) Combining Micro and Macro Data in Microeconometric Models, Review of Economic Studies, 61:4:655-680.

[58] Jarque Carlos M. and Anil K. Bera A Test for Normality of Observations and Regression Residuals: International Statistical Review 55, 2 (Aug., 1987), pp. 163-172

[59] Johansen Soren (1988) Statistical analysis of cointegration vectors, Journal of Economic Dynamics and Control, 12:231-254, North Holland.

[60] Johansen Soren (1988) Estimation and Hypothesis Testing of Cointegration Verctors in Gaussian Vector Autoregressive Models, Econometrica, 59:6, 1551-1580.

[61] Kadane Joseph B. and T. W. Anderson (1977) A Comment on the Test of Overidentifying Restrictions Econometrica, 45, 4 (May, 1977), pp. 1027-1031

[62] Keifer N (1988) Economic duration data and hazard functions, Journal of Economic Literature, 26:647-679.

[63] Klein L. R. and Mitsugu Nakamura (1962) Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicollinearity International Economic Review, 3, 3 (Sep., 1962), pp. 274-299

[64] King Maxwell L. (1981) The Durbin-Watson Test for Serial Correlation: Bounds for Regressions with Trend and/or Seasonal Dummy Variables Econometrica, 49, 6 (Nov., 1981), pp. 1571-1581

[65] Kiviet Jan F. (1986) On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships The Review of Economic Studies, 53, 2 (Apr., 1986), pp. 241-261

[66] Knight John L. (1986) Non-Normal Errors and the Distribution of OLS and 2SLS Structural Estimators Econometric Theory, 2, 1 (Apr., 1986), pp. 75-106

[67] Koop G (2003) Bayesian Econometrics, John Wiley and Sons

[68] Koop, Gary. (2008) Introduction to econometrics,Chichester : Wiley.

[69] Kumar T. Krishna (1975) Multicollinearity in Regression Analysis The Review of Economics and Statistics, 57, 3 (Aug., 1975), pp. 365-366

[70] Lancaster T (1979) Econometric Methods for Duration of Unemployment, Econometrica, 47:4:939-56.

[71] Lancaster T and A Chesher (1983) The Estimation of Models of Labour Market Behviour Review of Economic Studies, 50:4:609-624.

[72] Lancaster T (2004) An Introduction to Modern Bayesian Econometrics, Blackwell

[73] Lovell Michael C. (1986) Tests of the Rational Expectations Hypothesis The American Economic Review, 76, 1 (Mar., 1986), pp. 110-124

[74] Maddala G. S. (1971) Generalized Least Squares with an Estimated Variance Covariance Matrix Econometrica, 39, 1 (Jan., 1971), pp. 23-33

[75] McFadden Daniel (1963) Constant Elasticity of Substitution Production Functions, Review of Economic Studies, 30, 2, 73-83

[76] McFadden Daniel and Paul A. Ruud (1994) Estimation by Simulation, Review of Economics and Statistics, 76, 4 , 591-608

[77] Mills Terence C., Gianluigi Pelloni, Athina Zervoyianni (1995) Unemployment Fluctuations in the United States: Further Tests of the Sectoral-Shifts Hypothesis, The Review of Economics and Statistics, 77, 2,294-304

[78] Nabeya Seiji (2001) Approximation to the Limiting Distribution of t- and F-Statistics in Testing for Seasonal Unit Roots Econometric Theory, 17, 4 (Aug., 2001), pp. 711-737

[79] Nelson Charles R. (1987) A Reappraisal of Recent Tests of the Permanent Income Hypothesis The Journal of Political Economy, 95, 3 (Jun., 1987), pp. 641-646

[80] Patterson K (2000) A Primar to Unit Root Test, Approach,MacMillan.

[81] Pedroni, P. (1999): Critical values for cointegration tests in heterogeneous panels with multiple regressors, Oxford Bulletin of Economics and Statistics, 61, Special Issue,653-670.

[82] Pesaran M. H. (1982) A Critique of the Proposed Tests of the Natural Rate-Rational Expectations Hypothesis The Economic Journal, 92, 367 (Sep., 1982), pp. 529-554

[83] Pesaran, M.H. and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels", Journal of Econometrics, 68, p.79-113

[84] Phillips Peter C. B. (2003) Laws and Limits of Econometrics, Economic Journal, 113, 486, Conference Papers, pp. C26-C52

[85] Phillips P.C.B. (1987) Time Series Regression with an Unit Root, Econometrica, 55, 2, 277-301.

[86] Phillips, P. C. B., and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," Biometrika, 75, 335–346.

[87] Ramsey J. B. (1969) Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis Journal of the Royal Statistical Society. Series B (Methodological), 31, 2 (1969), pp. 350-371

[88] Rao Radhakrishna C (1992) R. A. Fisher: The Founder of Modern Statistics Statistical Science, 7, 1 (Feb., 1992), pp. 34-48

[89] Rao C R (1972) Recent Trends of Research Work in Multivariate Analysis, Biometrics- Special Multivariate Issue, 28:1:3-22.

[90] Rayner A C (1970) The Use of Multivariate Analysis in Development Theory: A Critique of the Approach Adopted by Adelman and Morris, Quarterly Journal of Economics, 84:4:639-647.

[91] Rider P R (1936) Annual Survey of Statistical Technique: Developments in the Analysis of Multivariate Data - Part I, Econometrica, 4:3:264-268.

[92] Schervish MJ (1987) A Review of Multivariate Analysis, Statistical Science, 2:4:396-413.

[93] Sim C.A (1980) Macroeconomics and Reality, Econometrica, 48:1 Jan.1 -45._

[94] Simon Julian L. and Dennis J. Aigner (1970) Cross-Section and Time-Series Tests of the Permanent-Income Hypothesis The American Economic Review, 60, 3 (Jun., 1970), pp. 341-351

[95] Sheshinski Eytan (1967) Tests of the "Learning by Doing" Hypothesis The Review of Economics and Statistics, 49, 4 (Nov., 1967), pp. 568-578

[96] Smith R.J, M. R. Weale, S. E. Satchell (1998) Measurement Error with Accounting Constraints: Point and Interval Estimation for Latent Data with an Application to U.K. Gross Domestic Product, The Review of Economic Studies, 65, 1, 109-134

[97] Staigler D., Stock J. H., (1997), "Instrumental Variables Regression with Weak Instruments", Econometrica, 65, 3, pp.557-586.

[98] Stock, J.H., and M.W. Watson (2001). Vector Autoregressions, Journal of Economic Perspectives, 15 (4): 101-115.

[99] Sorensen CK and JMP Gutierrez (2006) Euro Area Banking Sector Integration Using Hierarchical Cluster Analysis Techniques, European Central Bank, Working Paper Series, 626, May.

[100] Stevens DL (1973) Financial Characteristics of Merged Firms: A Multivariate Analysis, Journal of Financial and Quantitative Analysis, 8:2:149-158.

[101] Thomas BAM (1961) Some Industrial Applications of Multivariate Analysis, Applied Statistics, 10:1:1-8.

[102] Suits Daniel B. (1984) Dummy Variables: Mechanics V. Interpretation The Review of Economics and Statistics, 66, 1 (Feb., 1984), pp. 177-180

[103] Theil H. (1956) On the Theory of Economic Policy The American Economic Review, 46, 2, Papers and Proceedings of the Sixty-eighth Annual Meeting of the American Economic Association (May, 1956), pp. 360-366

[104] Theil H. (1965) The Analysis of Disturbances in Regression Analysis Journal of the American Statistical Association, 60, 312 (Dec., 1965), pp. 1067- 1079

[105] Wallace T. D. and C. E. Toro-Vizcarrondo (1969) Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression Journal of the American Statistical Association, 64, 328 (Dec., 1969), pp. 1649- 1663

[106] White Betsy Buttrill (1978) Empirical Tests of the Life Cycle Hypothesis The American Economic Review, 68, 4 (Sep., 1978), pp. 547-560

[107] White Halbert (9187) Consistency of OLS Econometric Theory, 3, 1,159-160

[108] Wooldridge Jeffrey M.(1994) Efficient Estimation under Heteroskedasticity Econometric Theory, 10, 1 (Mar., 1994), p. 223

[109] Zellner A. (1985) Bayesian Econometrics, Econometrica, 53:2:253-270

17.2 Applied Econometrics

References

[1] McGuinness Tony (1980) Econometric Analysis of Total Demand For Alcoholic Beverages in the U.K., 1956-75 The Journal of Industrial Economics, 29, 1 (Sep., 1980), pp. 85-109

[2] Shafik Nemat (1994) Economic Development and Environmental Quality: An Econometric Analysis Oxford Economic Papers, New Series, 46, Special Issue on Environmental Economics (Oct., 1994), pp. 757-773

[3] Lee Ray-Shine and Nirvikar Singh (1994) Patterns in Residential Gas and Electricity Consumption: An Econometric Analysis Journal of Business & Economic Statistics, 12, 2 (Apr., 1994), pp. 233-241

[4] Geweke John and William McCausland (2001) Bayesian Specification Analysis in Econometrics American Journal of Agricultural Economics, 83, 5, Proceedings Issue (Dec., 2001), pp. 1181-1186

[5] Carey Kathleen (2000) Hospital Cost Containment and Length of Stay: An Econometric Analysis Southern Economic Journal, 67, 2, 363-380

[6] Deolalikar Anil B. and Robert E. Evenson (1989) Technology Production and Technology Purchase in Indian Industry: An Econometric Analysis The Review of Economics and Statistics, 71, 4, 687-692

[7] Tony Lancaster and Andrew Chesher (1983) An Econometric Analysis of Reservation Wages Econometrica, 51, 6 (1661-1676

[8] Sims Christopher A., Stephen M. Goldfeld, Jeffrey D. Sachs (1982) Policy Analysis with Econometric Models Brookings Papers on Economic Activity, 1982, 1 (1982), pp. 107-164

[9] Van Soest Arthur and Peter Kooreman (1987) A Micro-Econometric Analysis of Vacation Behaviour Journal of Applied Econometrics, 2, 3 (Jul., 1987), pp. 215-226.

17.3 Textbooks in Econometrics

References

[1] Angrist, J.D. and Pischke, J.S., 2014. Mastering'metrics: The path from cause to effect. Princeton university press.

[2] Angrist, J.D. and Pischke, J.S., 2008. Mostly harmless econometrics. Princeton university press.

[3] Asteriou D. (2006) Applied Econometrics, Palgrave.

[4] Amemiya T. (1985) Advanced Econometrics, Harvard University Press.

[5] Baltagi H. (2008) Econometric analysis of panel data, Fourth Ediction, Blackwell.

[6] Bauwens L., M. Lubrano and J. F. Richard (1999) Bayesian Inference in Dynamic Econometric Models, Oxford.

[7] Bhattarai K. (2015) Econometric Analysis, University of Hull Business School.

[8] Blundell R. W.K. Newey and T. Persson (2006) Advances in Economics and Econometrics, Econometric Society Monograph, II and III, Cambridge University Press.

[9] Brooks C. (2002) Introductory Econometrics for Finance, Cambridge University Press.

[10] Burke S. P. and J. Hunter (2005) Modelling non-stationary economic time series, Palgrave.

[11] Campbell J. Y., A. W. Lo and A C MacKindlay (1997) The Econometrics of Financial Markets, Princeton.

[12] Chipman J. S. (2011) Advanced Econometric Theory, Routledge, London.

[13] Davidson J. (2000) Econometric Theory, Basil Blackwell Publishers.

[14] Davidson R and MacKinnon J. G. (2004) Econometric Theory and Methods, Oxford.

[15] Doornik J A and D.F. Hendry ((2012) PC-Give Volume I-III, GiveWin Timberlake Consultants Limited, London.

[16] Dougherty C. (2002) Introduction to Econometrics (Second Edition), Oxford University Press, 2002

[17] Enders W. (1995) Applied Econometric Time Series, John Wiley.

[18] GAMS Users Manual, GAMS Development Corporation, Washington DC. www.gams.com.

[19] Greene W. (2003) Econometric Analysis, Fifth Edition, Prentice Hall.

[20] George G. J, W.E. Griffiths, R.C. Hill, H. Lutkepohl T.C. Lee (1985) Theory and Practice of Econometrics, JohnWiley and Sons.

[21] Gujarati (G), D. N. (2005), Basic Econometrics, 4rd edition, McGraw Hill.

[22] Griffiths W. E., R.C. Hill and G. G. Judge, (1993), Learning and Practising Econometrics, J Wiley and Sons.

[23] Harvey A.C. (1989) Forecasting, structural time series and the Kalman filter, Cambridge University Press.

[24] Harvey A.C. (1990) The Econometric Analysis of Time Series, Phillip Allan, 2nd edition.

[25] Hendry D.F. (1995) Dynamic Econometric Theory, Oxford.

[26] Hendry D.F. and B. Nielsen (2007)Econometric Modelling: A Likelihood Approach, Princeton University Press.

[27] Hamilton J. D. (1994) Time Series Analysis, Princeton.

[28] Harris R. and R. Sollis (2003) Applied Time Series Modelling and Forecasting, John Wiley and Sons.

[29] Hill, C.R., W.E.Griffiths and G.C. Lim (2008), Principles of Econometrics, 3nd edition, John Wiley and Sons Inc. New York.

[30] Hsiao C. (1993) Analysis of panel data, Cambridge.

[31] Imbens, G., and D. Rubin (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

[32] Johnston J. (1960) Econometric methods, 7th edition, McGraw Hill.

[33] Koop G and D. Korobilis (2010) Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, now Publishers Inc., Hanover MA, USA

[34] Koop G, Dale J. Poirier and Justin L. Tobias (2010) Bayesian Econometric Methods, Cambridge University Press.

[35] Koop G. (2009) Analysis of Economic Data, Wiley, 2009.

[36] Kreps D. M. and K. F. Wallis (1997) Advances in Economics and Econometrics, Econometric Society Monograph, Vol. I Cambridge University Press.

[37] Kuthberson K. S. Hall and M. Taylor (1992) Applied time series analysis , Michigan University Press.

[38] Lancaster T. (1990) Econometric Analysis of Transition Data, Blackwell

[39] Lancaster T. (2004) An Introduction to Modern Bayesian Econometrics, Blackwell

[40] Maddala G.S. (2003) Introduction to Econometrics, Cambridge.

[41] Maddala G.S. (1983) Limited Dependent and Qualitative Variables in Econometrics, Cambridge.

[42] Patterson K (2000) An Introduction to Applied Econometrics: A time series approach, McMillan

[43] Pesaran M.H. (2015) Time series and panel data econometrics, Oxford University Press.

[44] Pyndick R. S. and D. L. Rubinfeld (1998) Econometric Models and Economic Forecasts, 4th edition, McGraw Hill

[45] Ruud P. A. (2000) An Introduction to Classical Econometric Theory, Oxford.

[46] Stock J. H. and M.W. Watson (2007) Introduction to Econometrics, 2nd edition, Addison Wesley.

[47] Syddeater K.P. Hammond P. and A. Seierstad and A. Strom (2008) Further Mathematics for Economic Analysis, second edition, Prentice Hall.

[48] Verbeek M. (2004) A Guide to Modern Econometrics, Wiley.

[49] K.F. Wallis et al. (1987) Models of the UK economy : a fourth review by the ESRC Macroeconomic Modelling Bureau , Oxford.

[50] Wang Peijie (2003) Financial Econometrics, Routledge Advanced Texts.

[51] Wooldridge J. M. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press.

[52] Wooldridge J.M. (2009) Introductory Econometrics: A Modern Approach, 4th ed., South Western.

17.4 Research in Economics in Hull

University Innovations Site :

- http://www2.hull.ac.uk/researchandinnovation.aspx
- http://www2.hull.ac.uk/researchandinnovation/researchareas.aspx

See Business School Research Papers from

- http://www2.hull.ac.uk/hubs/research.aspx

References

[1] Abbott Andrew , Joshy Easaw, Tao Xing (2008) Trade Integration and Business Cycle Convergence: Is the Relation Robust across Time and Space? Scandinavian Journal of Economics, 110, 2 , pp. 403-417

[2] Allanson P. and J P. Atkins (2005) The Evolution of the Racial Wage Hierarchy in Post-Apartheid South Africa, The Journal of Development Studies, 41, 6, August 2005, pp.1023 –1050

[3] Audas Rick, Tim Barmby, and John Treble (2004) Luck, Effort, and Reward in an Organizational Hierarchy,Journal of Labor Economics, 22, 2 , 379-395

[4] Basu P. and K. Bhattarai (2012) "Government Bias in Education, Schooling Attainment and Long-run Growth," *Southern Economic Journal*, 79(1) 127-142.

[5] Bandyopadhyay Taradas (1988) Revealed Preference Theory, Ordering and the Axiom of Sequential Path Independence, Review of Economic Studies, 55, 2, 343-351

[6] Barmby Tim, John Sessions and John Treble (1994) Absenteeism, Efficiency Wages and Shirking The Scandinavian Journal of Economics, 96, 4 , 561-566

[7] Bhattarai K and J. Whalley (2009) Redistribution Effects of Transfers, *Economica* 76:3:413-431 July.

[8] Bhattarai K. and J. Whalley (2003) Discreteness and the Welfare Cost of Labour Supply Tax Distortions, *International Economic Review* 44:3:1117-1133, August.

[9] Biswas T and Jolian McHardy(2005) Measuring the balance of intra-regional migration, Applied Economics, 2005, 37, 2221–2230

[10] Chesher A. and T. Lancaster (1983) The Estimation of Models of Labour Market Behaviour, *Review of Economic Studies*, 50, 4., Oct., 609-624.

[11] Dobson S. and C. Ramlogan (2012) Why is corruption less harmful to income inequality in Latin America?" World Development 40,8,1534-1545

[12] Dobson S. and C. Ramlogan (2012) Inequality, corruption and the informal sector, Economics Letters 115,1,104-107.

[13] Dolton P.J. and G. H. Makepeace (1990) The Earnings of Economics Graduates Economic Journal, 100, 399,. 237-250

[14] Dolton Peter and Donal O'Neill (1996) Unemployment Duration and the Restart Effect: Some Experimental Evidence The Economic Journal, 106, , 387-400

[15] Foreman-Peck James S. and Christopher J. Hammond (1997) Variable Costs and the Visible Hand: The Re-Regulation of Electricity Supply, 1932-37 Economica, , 64,. 253, 15-30

[16] Ghosh Sugata, Andros Gregoriou (2008) The Composition of Government Spending and Growth: Is Current or Capital Spending Better? Oxford Economic Papers, 60, 3 , 484-516

[17] Hammond C. J. (1991) Economic Computing in the 1990s; Recent Developments in Personal Computer HardwareThe Economic Journal, 101, 409 1611-1623

[18] Hammond C. J. (2006) Factor Substitution in the Production of Library Services: Evidence from the North American Research Libraries, Managerial and Decision Economics, 27, 8 , 613-630.

[19] Lancaster T (1979) Econometric Methods for Duration of Unemployment, *Econometrica*, 47:4:939-56.

[20] McCombie John S. L. (1987-88) Keynes and the Nature of Involuntary Unemployment, Journal of Post Keynesian Economics, 10, 2 , 202-215

[21] Mills Terence C. , Gianluigi Pelloni and Athina Zervoyianni (1995) Unemployment Fluctuations in the United States: Further Tests of the Sectoral-Shifts,Review of Economics and Statistics, 77, 2, 294-304

[22] Pike Maureen (1983) A Simulation of the Impact of Introducing Equality for Males and Females into Social Security, Economica, 50, 199, 351-360

[23] Ryan M. (2000) Economies of scale and scope, Manchester School. 68, 6, 701-722.

[24] Wilson J. S. G. (1949) Investment in a Monetary Economy, Economica, 16, 64, 321-335

[25] Terry Williams, Colin Eden, Fran Ackermann and Andrew Tait (1995) The Effects of Design Changes and Delays on Project Costs, The Journal of the Operational Research Society, 46, 7, 809-818

[26] Treble John G. and Simon Vicary (1993) Equity, Efficiency and Insurance: Explaining the Structure of Miners' Wage Payments in Victorian County Durham, Economic Journal, 103, 417, 481-493

[27] Turner Michael (1984) Agricultural Productivity in Eighteenth-Century England: Further Strains of Speculation, The Economic History Review, New Series, 37, 2, 252-257

17.5 Class test 2011

Answer any two questions; one from each section. Each question is worth 100 marks. Each sub-question has equal value in any question.

Section A

Q1. Consider y as a function of x_1, x_2 and x_3 as given in the following equation:

$$y = -5x_1^2 + 10x_1 + x_1x_2 - 2x_2^2 + 4x_2 + 2x_2x_3 - 4x_3^2 \tag{1866}$$

a. Find the optimal values of , and using the first order conditions for unconstrained maximisation. Use matrix approach in your solution.

b. Determine whether the above solutions correspond to the minimum or the maximum point using positive or negative definite concepts of the Hessian determinants.

Q2. Consider coefficients of a market model given by a matrix $A = \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}$

a) What are the eigen values of this maxtrix?

b) What are associated eigen vectors?

c) Prove that eigen vectors are orthogonal, $(V_1)^{'}(V_2) = 0$

d) Prove $(V_1V_2)^{'}(V_1V_2) = (V_1V_2)(V_1V_2)^{'} = I$

Section B

Q3. Consider the maximum likelihood function given below:

$$\text{ML:}\ \ lnL(\theta/Y) = \ln\left\{\prod_{i=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(Y-\beta X)^2}{\sigma^2}\right]\right\} \tag{1867}$$

a. How are parameters α, β, σ^2 estimated in this model?

b. For a linear function , prove that ML estimators of α, β, σ^2 are equivalent to the OLS estimators.

c. Discuss differences between the likelihood ratio test, Lagrange multiplier test and the Wald test. Use diagrams and equations to illustrate your answer.

d. Illustrate how the maximum likelihood method can be applied in estimating parameters in a ARMA(1,1) or ARCH(p,q) or GARCH(p,q) model.

Q4. Consider a structural VAR model between y_t and z_t as following:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1868}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1869}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$.

(a) Derive the reduced form of this VAR model and suggest ways to estimate the parameters in it.

(b) How would one determine stability of a VAR system like this? Provide analytical solutions using the roots of the quadratic function.

(c) How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified?

(d) Write impulse response functions for these two equations and indicate how can one perform an impulse response analysis with them?

(e) What is the meaning of variance decomposition in a VAR model like this?

Q5. Consider a monthly time series $\{y_t\}$

a. Show how the traditional moving average based methods could be applied to decompose its trend, seasonal, cyclical and irregular components.

b. Consider a random walk model $y_t = y_{t-1} + \varepsilon_t$ with initial conditions $y_1 = y_0$ for t =1.What are the mean, variance and the time path of y_t in terms of current and past series of errors ε_t ? What is its conditional forecast for period j made at time t? What is the error of forecast and its variance? How are the mean and variances affected if this random walk includes a drift term a_0 as in $y_t = y_{t-1} + a_0 + \varepsilon_t$.

c. Consider signal extraction problem for series y_t including permanent and transitory shocks components as ε_t and η_t

$y_t = \varepsilon_t + \eta_t$ and $\varepsilon_t^* = a + by_t$

where $E(\varepsilon_t) = 0$; $E(\eta_t) = 0$;$E(\varepsilon_t\eta_t) = 0$; $E\left(\varepsilon_t^2\right) = \sigma^2$; $E\left(\eta_t^2\right) = \sigma^2$.

What is its minimum square error (MSE)? How is the partitioning parameters b optimally estimated?

d. What are the prominent reasons for a failure of forecast? Illustrate Ganger and Newbold (1986) technique for combining optimal forecasts as in $f_{ct} = (1-\lambda) f_{1t} + \lambda f_{2t}$.

Q6. Write short notes in any three of the following.

a. Order of integration and unit root test.
b. Engle-Granger Representation theorem.
c. Johansen test for cointegrating vector.
d. Simultaneity bias.
e. Autocorrelation.

Q7. Consider a market model for a particular product.
Demand: $Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t}$ (1)
Supply: $Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t}$ (2)
Here Q_t^d is quantity demanded and Q_t^s is quantity supplied, P_t is the price of commodity,P_{t-1} is price lagged by one period, I_t is income of an individual, $u_{1,t}$ and $u_{2,t}$ are independently and identically distributed (iid) error terms with a zero mean and a constant variance.Q_t and P_t are endogenous variables and $P_{t-1}$and I_t are exogenous variables α_0,α_1,α_2,and β_0,β_1,$\beta_2$are six parameters defining the system.

1. How can simultaneity bias occur if one tries to apply OLS to each of the above equations.
2. Use rank and order conditions to judge whether each of these two equations are over-, under- or exactly identified.
3. Write down the reduced form for this system.
4. How would you estimate the coefficients of the reduced form equations? Write down the estimator.
5. If equations are identified explain how you may retrieve the structural parameters α_0,α_1,α_2,and β_0,β_1,β_2, and from the coefficients of the reduced form equations.

17.6 Class test 2012

Answer one question from each section

Section A

Q1. Consider a market model for a particular product.

Demand:

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t} \tag{1870}$$

Supply:

$$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t} \tag{1871}$$

Here Q_t^d is quantity demanded and Q_t^s is quantity supplied, P_t is the price of commodity,P_{t-1} is price lagged by one period, I_t is income of an individual, $u_{1,t}$ and $u_{2,t}$ are independently and identically distributed (iid) error terms with a zero mean and a constant variance.Q_t and P_t are endogenous variables and $P_{t-1}$and I_t are exogenous variables α_0,α_1,α_2,and β_0,β_1,$\beta_2$are six parameters defining the system.

a. How can a simultaneity bias occur if one tries to apply the OLS method to each of the above equations?

b. Use rank and order conditions to judge whether each of these two equations are over, under or exactly identified.

c. Write the reduced form for this system and define each of the reduced form coefficients in terms of structural parameters.

d. How would you estimate the coefficients of the reduced form equations? Write the estimator (s).

e. If equations are identified, explain how can one retrieve the structural parameters α_0,α_1,α_2,and β_0,β_1,$\beta_2$from the coefficients of the reduced form equations.

Q2. Consider a dynamic model of Y_t on Y_{t-1}, X_t and X_{t-1} in the **LSE** tradition as given below.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \tag{1872}$$

a. What is the error correction form (ECM) of this model?

b. Analyse implications of following restrictions in the comparative static or dynamic properties of this model.

1) $\alpha_1 = 0$; $\beta_1 = 0$
2) $\beta_1 = 0$
3) $\alpha_1 = 0$; $\beta_0 = 0$
4) $\alpha_1 = 0$
5) $\alpha_1 = 1$; $\beta_0 + \beta_1 = 0$
6) $\beta_0 = 0$

c. Assume the data generating process of X_t variable for each t follows a discrete Bernoulli probability likelihood function as

$$X(N) = p^{N_1}(1-p)^{N-N_1} \tag{1873}$$

X equals 1 if a draw happens to be a red ball and 0 otherwise. Let N_1 denote the number red balls out of total of N balls. Using the maximum likelihood estimator i) prove that the probability of X being red ($X = 1$) is $p = \frac{N_1}{N}$ and ii) it maximises the likelihood function (second order condition).

Section B

Q3. What is the main principle of forecasting and what are the reasons for failure of model based forecasts? Derive period 1 ahead forecast errors and variance of the forecast for each of the forecasting models given below.

a. A two period ahead ($h = 2$) forecast for a random walk with a drift: $[y_1 = y_0 + a_0 + \epsilon_1]$; $e_{T+1} \sim N(0,1)$

b. Period one ahead forecast ($h = 1$) of AR(1): $\left[y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h},\ e_{T+h} \sim N(0,1)\right]$.

c. Two period ahead ($h = 2$) forecast in MA(1): $\left[y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T,\ e_{T+1} \sim N(0,1)\right]$.

d. Two period ahead forecast ($h = 2$) in ARMA(1,1): $\left[y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1}\right]$; $e_{T+2} \sim N(0,1)$.

Q4. Consider a structural VAR model between y_t and z_t as following:

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \epsilon_{yt} \tag{1874}$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1875}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$.

(a) Derive the reduced form of this VAR model and suggest ways to estimate those reduced form parameters in it.

(b) How would one determine the stability of a VAR system like this? Provide analytical solutions using the roots of the relevant quadratic function.

(c) How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified?

(d) Write impulse response functions for these two equations and indicate how one could perform an impulse response to an unit shock in either of these two variables?

Q5. Write short notes in any three of the following.

(a) Engle-Granger Representation theorem.
(b) Johansen test for cointegrating vector.
(c) Nonparametric estimation
(d) Autocorrelation.
(e) Order of integration and unit root test.

17.7 Class test 2014

17.7.1 Section A

Q1. Consider a market model for a particular product where quantity demanded $\left(Q_t^d\right)$ and supplied (Q_t^s) are expressed as functions of the current and lagged prices, P_t and P_{t-1} of this commodity and the income of individuals (I_t) as:

Demand:

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_{1,t} \tag{1876}$$

Supply:

$$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2,t} \tag{1877}$$

here $u_{1,t}$ and $u_{2,t}$ are error terms representing missing elements of demand and supply functions. These are independently and identically distributed (iid) with zero means, constant variances and zero covariane: $u_{1,t} \sim N\left(0, \sigma_1^2\right)$; $u_{2,t} \sim N\left(0, \sigma_2^2\right)$ and $E\left(u_{1,t}u_{2,t}\right) = 0$. The quantity and price, Q_t and P_t, are endogenous variables and $P_{t-1}$and I_t are exogenous variables in this model. Then α_0,α_1,α_2, β_0, β_1 and β_2 are six unknown parameters of the system to be estimated from available data.

a. How can the simultaneity bias occur if the OLS is applied to each of the above equations?

b. Determine whether each of these two equations is over, under or exactly identified using the order and rank conditions.

c. Write the reduced form equations for this system and define each of the reduced form coefficients in terms of the structural parameters.

d. How would you estimate the coefficients in the reduced form equations? Characterise estimators that you suggest.

e. Dicsuss the procedure to retrieve the structural parameters $\widehat{\alpha}_0$,$\widehat{\alpha}_1$,$\widehat{\alpha}_2$, $\widehat{\beta}_0$,$\widehat{\beta}_1$and $\widehat{\beta}_2$ from the coefficients of the reduced form equations.

Q2. Consider a dynamic model of Y_t on Y_{t-1}, X_t and X_{t-1} in the **LSE** tradition as given below.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \tag{1878}$$

a. Express the error correction mechanism (ECM) for endogenous variable Y_t towards its long run equilibrium.

b. Analyse the implications of the following restrictions in the static or dynamic properties of this model.

  (a) $\alpha_1 = 0$; $\beta_1 = 0$
  (b) $\beta_1 = 0$
  (c) $\alpha_1 = 0$; $\beta_0 = 0$
  (d) $\alpha_1 = 0$
  (e) $\alpha_1 = 1$; $\beta_0 + \beta_1 = 0$
  (f) $\beta_0 = 0$

c. Discuss any three methods that can be applied to estimate unknown parameters $\alpha_0, \alpha_1, \beta_0$ and β_1 of this model. Which one of these methods is the best for ensuring the unbiasedness, efficiency and consistency properties of estimated parameters.

Q3. Consider a data generating process for a monthly macroeconomic or financial time series y_t, such as the unemployment rate, inflation, stock price or the exchange rate.

a. Show how the traditional moving average based methods can be applied to find the trend, seasonal, cyclical and irregular components of series y_t.

b. Assume that y_t follows a random walk, $y_t = y_{t-1} + \varepsilon_t$ with initial conditions $y_1 = y_0$ when $t = 1$. What are the mean, variance and the time path of y_t in terms of current and past series of errors ε_{t-s} for $s = 1, ., S$? How do the mean and variance of y_t change if this random walk process also includes a drift term a_0 as in it: $y_t = y_{t-1} + a_0 + \varepsilon_t$?

c. What is the main principle of forecasting and what are the reasons for the failure of model based forecasts?

d. Express the forecast errors for a

   (a) two period ahead ($h = 2$) forecast for a random walk with a drift: $[y_1 = y_0 + a_0 + \epsilon_1]$, $e_{T+1} \sim N(0, 1)$.
   (b) one period ahead forecast ($h = 1$) for AR(1) : $\left[y_{T+h} = \delta + \theta_1 y_{T+h-1} + e_{T+h},\ e_{T+h} \sim N(0, 1)\right]$.
   (c) two period ahead ($h = 2$) forecast for MA(1) : $\left[y_{T+1} = \mu + e_{T+1} + \alpha_1 e_T,\ \ e_{T+1} \sim N(0, 1)\right]$.
   (d) two period ahead forecast ($h = 2$) in ARMA(1,1) : $\left[y_{T+2} = \delta + \theta_1 y_{t+1} + e_{T+2} + \alpha_1 e_{T+1}\right]$; $e_{T+2} \sim N(0, 1)$.

17.7.2 Section B

Q4. Consider a structural VAR model between y_t and z_t:

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \epsilon_{yt} \tag{1879}$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1880}$$

where $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$.

a. Derive the reduced form equations of the VAR model above and suggest ways to estimate the reduced form parameters.

b. How would one determine the stability of a VAR system such as this? Base your answer on relevant quadratic roots emerging from your derivation.

c. How should one determine whether a VAR system as above is identified or not? What type of restrictions make it exactly or over identified?

d. Write the impulse response functions for y_t and z_t equations above and indicate how one could perform an impulse response to a unit shocks ϵ_{yt} or ϵ_{zt}?

e. Discuss the implications of the Engle-Granger Representation theorem and the Johansen's cointegrating test for for a VAR model as above.

Q5. Consider a panel data set with time $t = 1, ., ., .T$ and individuals $i = 1, ., ., .N$.

a. Discuss estimation procedures for the following types of panel data models:

   (a) Pooling cross section and time series in the SUR model.
   (b) Fixed effects model: $y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t} \qquad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$.

(c) Random effects model: $y_{i,t} = \alpha_i + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \qquad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$.

(d) Dynamic panel data model: $y_{i,t} = \alpha_i + \theta_i y_{i,t-1} + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$.

b. Explain the significance of following tests for a panel data model

(a) Im-Pesaran-Shin and KPSS panel unit root tests.

(b) Pedroni's panel cointegration test.

Q6. Explain the maximum likelihood estimation process for the cross section models given below.

(a) Logit $\quad P_i = \frac{1}{1+e^{-Z_i}}$ with $Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i$

(b) Count data model $\quad P\left(Y = y\right) = \frac{e^{-\lambda}\lambda^y}{y!}$

(c) Multinomial choice model: $\frac{P_{i,2}}{P_{i,1}} = \left(\left(\frac{\exp\left(X_{i,2}' \ \beta\right)}{\sum_{j=1}^{J}\exp\left(X_{i,j}' \ \beta\right)}\right) / \left(\frac{\exp\left(X_{i,1}' \ \beta\right)}{\sum_{j=1}^{J}\exp\left(X_{i,j}' \ \beta\right)}\right)\right) = \frac{\exp\left(X_{i,2}' \ \beta\right)}{\exp\left(X_{i,1}' \ \beta\right)}$.

(d) Explain Heckman's procedure to correct the selectivity bias in the cross section models as above.

17.8 Final exam 2011

Answer any four questions. Each question is of equal value. Each sub-question has equal value in any question.

Q1. Consider the Engle-Granger representation theorem.

a. Prove that there exists a valid error correction representation (ECM) of the data if two time series are cointegrated of order 1, $CI(1,1)$ with the following model.

$$Y_t = \varphi_2 X_t + \epsilon_t \tag{1881}$$

$$Y_t = X_t + \epsilon_t \ ; \qquad \varphi_2 = 1 \tag{1882}$$

$$\epsilon_t \ = Y_t - X_t \tag{1883}$$

b. Now generalise it to a vector error correction model (VECM) of the form

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1}\Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1884}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_3 + ... + \Pi_i \quad$ for $i = 1..k \quad \Gamma_k$ gives the long run solution and $\varepsilon_t \sim N(0, \sigma_y^2)$.

Prove that there exists at least one cointegrating vector in this system. Explain the Johansen procedure to determine the rank of the cointegrating vectors based on significant eigenvalues of the canonical correlations among the residuals of the system.

c. Consider the following model with $I(1)$ variables with $Y_{1,t}$ and $Y_{2,t}$ (but without exogenous X)

$$\Delta Y_{1,t} = -\frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \epsilon_{1,t} \tag{1885}$$

$$\Delta Y_{2,t} = \frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \epsilon_{2,t} \tag{1886}$$

$\Delta Y_{1,t} = \Delta Y_{2,t} = 0$ in the steady state; $E(\epsilon_{1,t}) = 0$; $E(\epsilon_{2,t}) = 0$. This implies the equilibrium relations $Y_{1,t-1} - \frac{1}{8}Y_{2,t-1} = 0$.

Illustrate how such an ECM model could be convrted in a VAR form.

Q2. Consider the Full Information Maximum Likelihood (FIML) method to estimate a standard macroeconomic model of the form

$$C = a_0 + a_1(Y - T) + a_2 Y_{t-1} + u_1 \tag{1887}$$

$$I = b_0 + b_1 Y + b_2 K_{t-1} + b_3 r + b_4 E + u_2 \tag{1888}$$

$$M = c_0 + c_1 Y + c_2 P_{t-1} + u_3 \tag{1889}$$

$$Y = C + I + G + E - M \tag{1890}$$

where the endogenous variables Y, C, I, M are the levels of income, consumption, investment and imports respectively; exogenous variables $G, E,\ T, K_{t-1}, P_{t-1}$ and r denote the government spending, exports, tax revenue, lagged capital, lagged price level and the interest rate respectively; terms $u_1, u_2$and u_3 denote errors in consumption, investment and import functions respectively. Parameters a_0, a_1, a_2, b_0, b_1, b_2, b_3, b_4, c_0, c_1 and c_2 provide the behavioral relations in the system.

a. Derive the reduced form coefficients of this system in terms of structural parameters a_0, a_1, a_2, b_0, b_1, b_2, b_3, b_4, c_0, c_1 and c_2.Write the reduced form equations in terms of redefined endogenous variables $y_1 = C$; $y_2 = I$; $y_3 = M$ and the redefined exogenous variables $z_1 = G; z_2 = E; z_3 = K_{t-1}; z_4 = T; z_5 = Y_{t-1}; z_6 = r; z_7 = P_{t-1}$ and the redefined reduced form parameters α_0, $\alpha_1, \alpha_2, \beta_0$, $\beta_1, \beta_2, \beta_3, \beta_4, \gamma_0$, γ_1 and γ_2.

b. Write expressions for the joint distribution of errors, $u_{11}u_{12}u_{13}......u_{1n}; u_{21}u_{22}u_{23}......u_{2n}; u_{31}u_{32}u_{33}......u_{3n}$, in this system.

c. Derive the Jacobian matrix of the first order conditions to estimate the parameters of the reduced system:

$$u_1 = y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5 \tag{1891}$$

$$u_2 = y_2 - \beta_0 - \beta_1(y_1) + \beta_1(y_3) - \beta_1(z_1) - \beta_2 z_3 - \beta_3 z_6 - \beta_4 z_2 \tag{1892}$$

$$u_3 = y_3 - \gamma_0 - \gamma_1(y_1) - \gamma_1(y_2) - \gamma_1(z_1) - \gamma_1(z_2) - \gamma_2 z_7 \tag{1893}$$

d. Derive the log likelihood functions for y variables using the transformation functions as:

$$P\left(y_{1i}.y_{2i}.y_{3i}\right) = P\left(u_{1i}.u_{2i}.u_{3i}\right).\left|\frac{\partial\left(u_{1i}.u_{2i}.u_{3i}\right)}{\partial\left(y_{1i}.y_{2i}.y_{3i}\right)}\right| \tag{1894}$$

e. Show the first order conditions for maximisation of the log likelihood function in terms of the reduced form parameters α_0, $\alpha_1, \alpha_2, \beta_0$, $\beta_1, \beta_2, \beta_3, \beta_4, \gamma_0$, γ_1 and γ_2.

f. Explain the procedure by which all FIML parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2$ and $\sigma^2_{u_1}, \sigma^2_{u_2}$ and $\sigma^2_{u_3}$ could be estimated to evaluate the log of FIML.

g. Why the FIML is the most efficient estimation technique for a system like this?

Q3. Consider a consumer optimisation problem in a capital asset pricing model popularised by Hansen and Singleton (1982) stated as:

$$max\ E_0\left[\sum_{t=0}^{\infty}\delta^t U\left(C_t\right)\right] \tag{1895}$$

subject to

$$C_t + P_t Q_t \leq R_t Q_{t-1} + W_t \tag{1896}$$

where C_t denotes consumption at time t, P_t the price level, Q_t assets, R_t return on assets and W_t wage income, U denotes utility, δ discount factor and E_0 is the expectation operator.

a. Formulate the generic Lagrangian function for constrained optimisation to solve this problem. Then specify the utility function as $U\left(C_t\right) = \frac{C_t^{\gamma}}{\gamma}$ with $\gamma < 1$; $\alpha = \gamma - 1$.

b. Derive the Euler equations based on the first order conditions that maximise the objective function of the model.

c. Discuss how the GMM could be applied to estimate the parameters δ and α. Why is the application of the maximum likelihood computationally challenging here? Why does it generate biased results?

d. Discuss properties of Arellano-Blundell-Bond GMM estimators for a dynamic panel data model to estimate such a CAPM model across countries.

e. Discuss the GMM estimation procedure in popular softwares such as Eviews/RATS/STATA/PcGive for applied studies.

Q4. Consider a data set with time $t = 1, .., .., .T$ and individuals $i = 1, \ldots .N$.

a. Discuss the specification, estimation and testing procedure for pooling cross section and time series in a SUR model.

b. Show how the α_i and β parameters can be estimated using the least square dummy variable panel data model in $y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma^2_{i,t}\right)$. Why is it called a fixed effect model?

c. What is the estimation and testing procedure for a random effect panel data model of the form $y_{i,t} = \alpha_i + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma^2_{i,t}\right)$?

d. Illustrate how the GMM is the most efficient estimation method for a dynamic panel data model of the form: $y_{i,t} = \alpha_i + \theta_i y_{i,t-1} + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma^2_{i,t}\right)$?

e. Explain the need for Im-Pesaran-Shin and KPSS panel unit root tests in panel data models.

f. What are the procedures for Pedroni's panel cointegration test?

Q5. Consider a standard regression model $Y = \beta X + e$ where Y is $T \times 1$ is the vector of dependent variable, X is $T \times K$ is the matrix of explanatory variables, e is $T \times 1$ the vector of independently and identically distributed normal random variables with the mean equal to zero and a constant variance, that is $e \sim N(0, \sigma^2 I)$. Here β is a $K \times 1$ vector of the unknown coefficients.

a. Show how best, linear and unbiased parameters $\theta = \left(\beta, \sigma^2\right)$ can be estimated using the OLS method.

b. What is exact multicollinearity? Show its consequences on estimation of vector of parameters, β. How could one estimate the variance inflation factor in case of inexact multicollinearity?

c. Show briefly how the generalised least square method can be used to avoid the heteroskedasticity or autocorrelation problems.

d. Discuss a procedure and the test statistics to test restrictions on the OLS estimator of parameter vector β.

e. Prove that the ML estimators of β, σ^2 are equivalent to the OLS estimators [ML: $lnL(\theta/Y) = \ln\left\{\prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{(Y-\beta X)^2}{\sigma^2}\right]\right\}$].

f. Show that the GMM estimators of β and σ^2 are equivalent to both the OLS and ML estimators.

Q6. Consider a structural VAR model between y_t and z_t with errors $\epsilon_{yt} \sim N(0, \sigma^2_y)$ and $\epsilon_{zt} \sim N(0, \sigma^2_z)$ as following:

$$y_t = b_{10} - b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \epsilon_{yt} \tag{1897}$$

$$z_t = b_{20} - b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \epsilon_{zt} \tag{1898}$$

a. Derive the reduced form of this VAR model and suggest ways to estimate the parameters in it.

b. How would one determine stability of a VAR system like this? Provide analytical solutions using the roots of the quadratic functions.

c. Write the impulse response functions for these two equations. Indicate how the impulse response analysis could be performed with them.

d. How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified? How could one retrieve the structural parameters?

e. What is the meaning of variance decomposition in a VAR model like this?

f. Generalise this VAR to the form $X_t = A_0 + A_1 X_{t-1} + \varepsilon_t$. Use undetermined coefficient model in order to ascertain the stability in it.

Q7. Consider modelling the probability (p_i) of occurrence of a particular economic event.

a. Contrast the probit, logit and linear probability models. Show how a change in an independent variable impacts on the probability of occurrence of such event. How are the probit and logit models superior to a linear probability model?

b. Illustrate how the parameters of the probit and logit models could be estimated using the maximum likelihood method. Explain the underlying Newton-Ralphson or BHHH algorithms.

c. Specify a multinomial logit model. How can the independence of irrelevant alternatives (IIA) assumption be used in making a choice between two alternatives among a set of j alternatives?

d. How do truncation and censoring affect the mean and variance of a random variable? Illustrate using diagrams and equations.

e. Specify a regression model for a truncated dependent variable. Show the Tobit two stage procedure for estimating a regression model with a truncated variable.

f. What is the meaning of selection bias in a non-random sample? How does Heckman's lambda correct for the selectivity bias? What is the procedure to estimate a wage determination model with selectivity bias?

g. "Correction of selectivity bias is as important in micro-econometrics as the unit root test is for time series modelling." Comment.

[Hints: Logit: $\left[P_i = \frac{1}{1+e^{-Z_i}} \text{ with } Z_i = \beta_1 + \beta_2 X_i + \varepsilon_i\right]$;Count data: $\left[P(Y=y) = \frac{e^{-\lambda}\lambda^y}{y!}\right]$;Ordered probit model: $\left[prob\,(y = J|x) = 1 - \Phi\left(\mu_{J-1} - x'\beta\right)\right]$;Multinomial Choice model:

$$\left[\frac{P_{i,2}}{P_{i,1}} = \frac{\frac{\exp\left(X'_{i,2}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}}{\frac{\exp\left(X'_{i,1}\ \beta\right)}{\sum_{j=1}^{J}\exp\left(X'_{i,j}\ \beta\right)}} = \frac{\exp\left(X'_{i,2}\ \beta\right)}{\exp\left(X'_{i,1}\ \beta\right)}\right].$$

Heckman's correction for selectivity bias in which $Y_{1,i} = X_{1i}\beta + \varepsilon_{1,i}$ and $Y_{2,i} = X_{2,i}\beta + \varepsilon_{2,i}$ and if the event occurs $Y_{1,i} > Y_{2,i}$.

Two limit Tobit for a certain regression $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ with

$$Y_i = \begin{pmatrix} Y_i^* \ \ if \ \ L_1 < \ Y_i^* < L_2 \text{ if the event occurs} \\ L_1 \ \ \text{if } Y_i^* < L_1 \\ L_2 \ \ \text{if } Y_i^* \geq L_2 \end{pmatrix}].$$

Q8. Consider each of the following duration or hazard function models:

- Exponential distribution. $[F(t) = 1 - \exp(-\lambda.t)]$.
- Weibull $\left[f(t) = \alpha\lambda t^{\alpha-1}\exp(-\lambda.t^{\alpha})\right]$.
- Log normal distribution $\left[f(t) = \frac{1}{\sigma.t}\phi\left(\frac{\log T - m}{\sigma}\right)\right]$.

- Log logistic $\left[f(t) = \alpha\gamma t^{\alpha-1}\left(1+\gamma t^{\alpha}\right)^{-2}\right]$.
- Gamma distribution $\left[f(t) = \frac{\left[a^{v}t^{v-1}\exp(-at)\right]}{\Gamma(v)} \quad \text{where} \quad \Gamma(v) = \int_0^{\infty}\exp(-t)\,t^{v-1}\partial s.\right]$

a. Derive the duration density, the hazard rate, the survival function and the duration dependence in each of these models.

b. Explain the general procedure for estimation of model parameters.

c. Discuss briefly issues for which models could be applied.

Q9. Provide answers to any five of the following questions with basic relevant derivations appropriate to the Bayesian modelling and analysis.

a. Difference between classical and Bayesian assumptions on parameters and errors in a linear regression.

b. Bayesian rule where $p(A, B)$ denotes the joint probability of occurring events A and B together, $p(B)$ is the marginal probability of B without any respect to occurrence of A. Similarly $p(A)$ is the marginal probability of A without any respect to occurrence of B.

c. Bayesian prior and posterior density functions on unknown parameters β and τ for the likelihood function such as: $\log l(\beta, \tau) = \frac{1}{2}\left[n\log\tau - \tau(\beta - b)X'X(\beta - b) - \tau e'e\right]$

d. Estimates of the mean and variance of β and τ in the above Bayesian linear regression model.

e. Bayesian panel data model of the form $y_{i,t} = x_{i,t}\beta + \alpha_i + e_{i,t}$ with priors $e_{i,t}/x_{i,t}, \beta, \alpha_i, \tau \sim N(0, \tau)$.

f. Estimation procedure in a Bayesian VAR model of the form: $\underset{T\times 1}{Y_j} = \underset{T\times K}{X_j}\ \underset{K\times 1}{\beta_j} + \underset{T\times 1}{u_j}$ where $u_j \sim N\left(0, \sigma_{i,j}^2 I_T\right)$.

g. Specification and estimation procedure of the Bayesian stochastic volatility models (Bayesian ARCH/GARCH).

h. MCMC algorithm.

i. Ito calculus

j. Brownian motion

17.9 Final exam 2014

Q1. Consider a version of the Hicksain formulation of Keynesian macroeconometric model in spirit of Klein (1947) with following six eqautions with income (Y_t), interest rate (R_t), consumption (C_t), investment (I_t), tax revenue (T_t) and imports (M_t) as endogenous variables. It includes government spending (G_t), exports (X_t), real money balances-liquidity $\left(\frac{\overline{MM}_t}{P_t}\right)$ as exogenous and policy variables

Consumption function:

$$C_t = \beta_0 + \beta_1 (Y_t - T_t) + \beta_2 X_t + \varepsilon_{1,t} \tag{1899}$$

Taxes:

$$T_{\mathbf{t}} = t_0 + t_1 Y_t + t_2 M_t + t_3 G_t + \varepsilon_{2,t} \tag{1900}$$

Imports:

$$M_{\mathbf{t}} = m_0 + m_1 Y_t + m_2 M_t + m_3 G_t + \varepsilon_{3,t} \tag{1901}$$

Investment:

$$I_{\mathbf{t}} = \mu_0 - \mu_1 R_t + \phi \Delta Y_{t-1} + \varepsilon_{4,t} \tag{1902}$$

Money Market:

$$\left(\frac{\overline{MM}_t}{P_t}\right)_t = b_0 + b_1 Y_t - b_2 R_t + \varepsilon_{5,t} \tag{1903}$$

Macro balance

$$Y_t = C_t + I_t + G_t + X_t - M_t = C_t + T_t + S_t \tag{1904}$$

Money Market Equilibrium

$$R_t = \frac{b_0}{b_2} - \frac{1}{b_2}\left(\frac{\overline{MM}_t}{P_t}\right) + \frac{b_1}{b_2} Y_t \tag{1905}$$

Here $\varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{5,t}$ are ideiosyncratic errors terms normally distributed with zero mean and constant variances; $\varepsilon_{i,t} \sim N\left(0, \sigma^2\right)$ for $i = 1, ..5$. Behavioral parameters are β_0, β_1 $T_0, \mu_0, m_0, t_1, m_1, b_0, b_1, b_2$ are unknown and need to be estimated using quarterly time series data on those variables.

1. Construct a table of structural coefficients in the tabular form.
2. Check whether each of the equation is identified or not using rank and order conditions.
3. Write the reduced form of the model.
4. Discuss any four methods that can be applied to estimate this model using quarterly time series data.
5. Show procedure for the historical simulations using this model and procedure to project the values of exogenous variables $\left(G_t, X_t, \frac{\overline{MM}_t}{P_t}\right)$ for the next 10 quarters.

6. How can one judge the accuracy of forecast in this economy?

[56268] [Continued...]

Q2. Consider a structural VAR model between y_t and z_t with errors $\epsilon_{yt} \sim N(0, \sigma_y^2)$ and $\epsilon_{zt} \sim N(0, \sigma_z^2)$ as follows:

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \epsilon_{yt} \tag{1906}$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{21}y_{t-1} + \gamma_{22}z_{t-1} + \epsilon_{zt} \tag{1907}$$

a. Derive the reduced form of this VAR model and suggest how the maximum likehoold estimator (MLE) could be applied to estimate the parameters in it.

b. How would one determine stability of a VAR system like this? Provide analytical solutions using the roots of the quadratic functions.

c. Write the impulse response functions for these two equations. Indicate how the impulse response analysis could be performed with them.

d. How should one determine whether a VAR system like this is identified or not? What sort of restrictions make it exactly or over identified? How could one retrieve the structural parameters?

e. What is the meaning of variance decomposition in a VAR model like this?

f. Generalise this VAR to the form $X_t = A_0 + A_1 X_{t-1} + \varepsilon_t$. Use undetermined coefficient model in order to ascertain the stability in it.

g. What additional insights are obtained by transforming this model into a vector error correction model (VECM)?

h. Why is the Bayesian VAR becoming more popular than a classical VAR as given above in recent years?

[56268] [Continued...]

Q3. Consider a standard OLS regression model $Y = \beta X + e$ where Y is $T \times 1$ vector of dependent variable, X is $T \times K$ matrix of explanatory variables, e is $T \times 1$ vector of independently and identically distributed normal random variable with mean equal to zero and a constant variance, that is $e \sim N(0, \sigma^2 I)$. Here β is a $K \times 1$ vector of unknown coefficients.

a. What is the estimator of the parameter vector $\left(\widehat{\beta}\right)$ in the unrestricted least squares method? Prove that the OLS provides the best, linear and unbiased estimators of the parameters $\theta = \left(\beta, \sigma^2\right)$ in this model.

b. Show briefly how the generalised least square method can be used to avoid heteroskedasticity or autocorrelation problems.

c. Consider a Lagrange function to minimise sum of errors $(e'e)$ for a restricted least squares model of Y on X with restrictions $\left(r' - \beta' R'\right) = 0$.

$$\begin{aligned} L &= e'e + 2\lambda\left(r' - \beta' R'\right) = (Y - \beta X)'(Y - \beta X) + 2\lambda\left(r' - \beta' R'\right) \\ &= Y'Y - 2\beta X'Y + (\beta X)'(\beta X) + 2\lambda\left(r' - \beta' R'\right) \end{aligned} \tag{1908}$$

Using the first order conditions derive the estimator of parameter vector $\left(\widehat{b}\right)$ given by the restricted least squares method. What is the value of λ?

d. Prove that the variance of the parameter vector $\left(\widehat{b}\right)$ in the restricted least squares model is smaller than the variance of the parameter vector $\left(\widehat{\beta}\right)$ in the unrestricted least squares model.

e. Prove that the maximul likelihood (ML) estimators of β, σ^2 are equivalent to the OLS estimators, where:

$$ML: \quad lnL(\theta/Y) = \ln\left\{\prod_{i=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2}\frac{(Y - \beta X)^2}{\sigma^2}\right]\right\} \tag{1909}$$

f. Show that the GMM estimators of β, σ^2 are equivalent to both the OLS and ML estimators.

g. Why are GMM and ML estimators more are popular in advanced studies than the OLS estimators? Comment with some examples.

[56268] [Continued...]

Q4. Consider the Engle-Granger representation theorem.

a. Prove that there exists a valid error correction representation (ECM) of the data if two time series are cointegrated of order 1, $CI(1,1)$ with the following model.

$$Y_t = \varphi_2 X_t + \epsilon_t \tag{1910}$$

$$Y_t = X_t + \epsilon_t \ ; \qquad \varphi_2 = 1 \tag{1911}$$

$$\epsilon_t \ = Y_t - X_t \tag{1912}$$

b. Now generalise it to a vector error correction model (VECM) of the form

$$\Delta X_t = \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \Gamma_k X_{t-k} + \varepsilon_t \tag{1913}$$

where $\Gamma_1 = -I + \Pi_1 + \Pi_2 + \Pi_3 + ... + \Pi_i \quad$ for $i = 1..k \quad \Gamma_k$ gives the long run solution and $\varepsilon_t \sim N(0, \sigma_y^2)$.

Prove that there exists at least one cointegrating vector in this system. Explain the Johansen procedure to determine the rank of the cointegrating vectors based on significant eigenvalues of the canonical correlations among the residuals of the system.

c. Consider the following model with $I(1)$ variables $Y_{1,t}$ and $Y_{2,t}$ (but without any exogenous variable X)

$$\Delta Y_{1,t} = -\frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \epsilon_{1,t} \tag{1914}$$

$$\Delta Y_{2,t} = \frac{1}{2}\left(Y_{1,t-1} - \frac{1}{8}Y_{2,t-1}\right) + \epsilon_{2,t} \tag{1915}$$

$\Delta Y_{1,t} = \Delta Y_{2,t} = 0$ in the steady state; $E\left(\epsilon_{1,t}\right) = 0$; $E\left(\epsilon_{2,t}\right) = 0$. This implies the equilibrium relations $Y_{1,t-1} - \frac{1}{8}Y_{2,t-1} = 0$.

Illustrate how such an ECM model could be convrted in a VAR form.

d. Discuss generic to specific modeling philosophy in VAR and ECM developed in Hendry (1995) .

[56268] [Continued...]

Q5. Consider the Full Information Maximum Likelihood (FIML) method to estimate a standard macroeconomic model of the form

$$C = a_0 + a_1(Y - T) + a_2 Y_{t-1} + u_1 \tag{1916}$$

$$I = b_0 + b_1 Y + b_2 K_{t-1} + b_3 r + b_4 E + u_2 \tag{1917}$$

$$M = c_0 + c_1 Y + c_2 P_{t-1} + u_3 \tag{1918}$$

$$Y = C + I + G + E - M \tag{1919}$$

where the endogenous variables Y, C, I, M are the levels of income, consumption, investment and imports respectively; exogenous variables $G, E,\ T, K_{t-1}, P_{t-1}$ and r denote the government spending, exports, tax revenue, lagged capital, lagged price level and the interest rate respectively; terms $u_1, u_2$and u_3 denote errors in consumption, investment and import functions respectively. Parameters $a_0,\ a_1, a_2, b_0,\ b_1, b_2, b_3, b_4, c_0,\ c_1$ and c_2 provide the behavioral relations in the system.

a. Derive the reduced form coefficients of this system in terms of structural parameters a_0, $a_1, a_2, b_0,\ b_1, b_2, b_3, b_4, c_0,\ c_1$ and c_2.Write the reduced form equations in terms of redefined endogenous variables $y_1 = C$; $y_2 = I$; $y_3 = M$ and the redefined exogenous variables $z_1 = G; z_2 = E; z_3 = K_{t-1}; z_4 = T; z_5 = Y_{t-1}; z_6 = r; z_7 = P_{t-1}$ and the redefined reduced form parameters α_0, $\alpha_1, \alpha_2, \beta_0,\ \beta_1, \beta_2, \beta_3, \beta_4, \gamma_0,\ \gamma_1$ and γ_2.

b. Write expressions for the joint distribution of errors, $u_{11}u_{12}u_{13}......u_{1n}; u_{21}u_{22}u_{23}......u_{2n}; u_{31}u_{32}u_{33}......u_{3n}$, in this system.

c. Derive the Jacobian matrix of the first order conditions to estimate the parameters of the reduced system:

$$u_1 = y_1 - \alpha_0 - \alpha_1(y_2) + \alpha_1(y_3) - \alpha_1(z_1) - \alpha_1(z_2) + \alpha_1(z_4) - \alpha_2 z_5 \tag{1920}$$

$$u_2 = y_2 - \beta_0 - \beta_1(y_1) + \beta_1(y_3) - \beta_1(z_1) - \beta_2 z_3 - \beta_3 z_6 - \beta_4 z_2 \tag{1921}$$

$$u_3 = y_3 - \gamma_0 - \gamma_1(y_1) - \gamma_1(y_2) - \gamma_1(z_1) - \gamma_1(z_2) - \gamma_2 z_7 \tag{1922}$$

d. Derive the log likelihood functions for y variables using the transformation functions as:

$$P(y_{1i}.y_{2i}.y_{3i}) = P(u_{1i}.u_{2i}.u_{3i}).\left|\frac{\partial(u_{1i}.u_{2i}.u_{3i})}{\partial(y_{1i}.y_{2i}.y_{3i})}\right| \tag{1923}$$

e. Show the first order conditions for maximisation of the log likelihood function in terms of the reduced form parameters $\alpha_0,\ \alpha_1, \alpha_2, \beta_0,\ \beta_1, \beta_2, \beta_3, \beta_4, \gamma_0,\ \gamma_1$ and γ_2.

f. Explain the procedure by which all FIML parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4, \gamma_1, \gamma_2$ and $\sigma^2_{u_1}, \sigma^2_{u_2}$and $\sigma^2_{u_3}$ could be estimated to evaluate the log of the FIML.

g. Why the FIML is the most efficient estimation technique for a system like this?

[56268] [Continued...]

Q6. Consider a consumer optimisation problem in a capital asset pricing model popularised by Hansen and Singleton (1982) stated as:

$$max\ E_0 \left[\sum_{t=0}^{\infty} \delta^t U\left(C_t\right) \right] \tag{1924}$$

subject to

$$C_t + P_t Q_t \leq R_t Q_{t-1} + W_t \tag{1925}$$

where C_t denotes consumption at time t, P_t the price level, Q_t assets, R_t return on assets and W_t wage income, U denotes utility, δ discount factor and E_0 is the expectation operator.

a. Formulate the generic Lagrangian function for constrained optimisation to solve this problem. Then specify the utility function as $U\left(C_t\right) = \frac{C_t^{\gamma}}{\gamma}$ with $\gamma < 1$; $\alpha = \gamma - 1$.

b. Derive the Euler equations based on the first order conditions that maximise the objective function of the model.

c. Discuss how the GMM could be applied to estimate the parameters δ and α. Why is the application of the maximum likelihood computationally challenging here? Why does it generate biased results?

d. Discuss properties of Arellano-Blundell-Bond GMM estimators for a dynamic panel data model to estimate such a CAPM model across countries.

e. Discuss the GMM estimation procedure in popular softwares such as Eviews/RATS/STATA/PcGive for applied studies.

Q7. Consider a data set with time $t = 1, .., .., .T$ and individuals $i = 1, \ldots .N$.

a. Discuss the specification, estimation and testing procedure for pooling cross section and time series in a SUR model.

b. Show how the α_i and β parameters can be estimated using the least square dummy variable panel data model in $y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$. Why is it called a fixed effect model?

c. What is the estimation and testing procedure for a random effect panel data model of the form $y_{i,t} = \alpha_i + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$?

d. Illustrate how the GMM is the most efficient estimation method for a dynamic panel data model of the form: $y_{i,t} = \alpha_i + \theta_i y_{i,t-1} + x_{i,t}\beta + \gamma_i + \varepsilon_{i,t} \quad \varepsilon_{i,t} \sim N\left(0, \sigma_{i,t}^2\right)$?

e. Explain the need for Im-Pesaran-Shin and KPSS panel unit root tests in panel data models.

f. What are the procedures for Pedroni's panel cointegration test?

18 Foundations

18.1 First order difference equation

Supply

$$Q_{d,t} = \alpha - \beta P_t \ ; \quad (\alpha, \beta > 0) \tag{1926}$$

Demand

$$Q_{S,t} = -\gamma + \delta P_{t-1}; \qquad (\gamma, \delta > 0) \tag{1927}$$

$$Q_{d,t} = Q_{S,t} \Longrightarrow \beta P_t + \delta P_{t-1} = \alpha + \gamma \tag{1928}$$

Steady state or intertemporal solution

$$\overline{P} = \frac{\alpha + \gamma}{\beta + \delta} \tag{1929}$$

Complete solution $P_t = P_C + P_P$
Complementary solution

$$\beta P_t + \delta P_{t-1} = 0 \Longrightarrow P_{t+1} + \frac{\delta}{\beta} P_t = 0 \tag{1930}$$

Let $P_t = Ab^t$ and $P_{t+1} = Ab^{t+1}$

$$Ab^{t+1} + \frac{\delta}{\beta} Ab^t = 0 \tag{1931}$$

Steady state or intertemporal solution

$$b = -\frac{\delta}{\beta} \tag{1932}$$

$$P_t = P_C + P_P = Ab^t + P_P = A\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} \tag{1933}$$

Determiner A from the initial condition P_0
$P_0 = Ab^0 + P_P = A\left(-\frac{\delta}{\beta}\right)^0 + \frac{\alpha+\gamma}{\beta+\delta} \Longrightarrow$

$$A = P_0 - \frac{\alpha + \gamma}{\beta + \delta} \tag{1934}$$

Complete and definite solution

$$P_t = A\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta}\right)\left(-\frac{\delta}{\beta}\right)^t + \frac{\alpha + \gamma}{\beta + \delta} \tag{1935}$$

Application of First Order Difference Equation: Inventory and Price Adjustment Model
Consider a demand supply model with inventory and price adjustments.
Demand depends on current price:

$$X_t^d = \alpha - \beta P_t \tag{1936}$$

Supply depends on current price:

$$X_t^S = -\gamma + \delta P_t \tag{1937}$$

Price adjustment process:

$$P_{t+1} = P_t + \sigma \left(X_t^d - X_t^S \right) \tag{1938}$$

Equilibrium conditions without inventory would be $X_t^d = X_t^S$ but here prices do not clear market instantly. Therefore prices adjust according to:

$$P_{t+1} = P_t + \sigma \left(\alpha - \beta P_t + \gamma - \delta P_t \right) \tag{1939}$$

This is a first order difference equation in prices.

Application of First Order Difference Equation: Inventory and Price Adjustment Model

$$P_{t+1} - \left(1 - \sigma \left(\beta + \delta\right)\right) P_t = \sigma \left(\alpha + \gamma\right) \tag{1940}$$

Intertemporal solution

$$\overline{P} = \frac{(\alpha + \gamma)}{(\beta + \delta)} \tag{1941}$$

Whether prices converge to this stationary solution depends on solutions to the complementary part

$P_{t+1} - \left(1 - \sigma \left(\beta + \delta\right)\right) P_t = 0$

$$Ab^{t+1} - \left(1 - \sigma \left(\beta + \delta\right)\right) Ab^t = 0 \tag{1942}$$

$b = \left(1 - \sigma \left(\beta + \delta\right)\right)$;

Application of First Order Difference Equation: Inventory and Price Adjustment Model

The general solution for price is:

$$P_t = P_C + P_P = A \left(\left(1 - \sigma \left(\beta + \delta\right)\right)\right)^t + \frac{(\alpha + \gamma)}{(\beta + \delta)} \tag{1943}$$

Value of A can be obtained by assuming initial price at time $t = 0$, P_0 implies $A = P_0 - \frac{\alpha+\gamma}{\beta+\delta}$

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta} \right) \left(1 - \sigma \left(\beta + \delta\right)\right)^t + \frac{(\alpha + \gamma)}{(\beta + \delta)} \tag{1944}$$

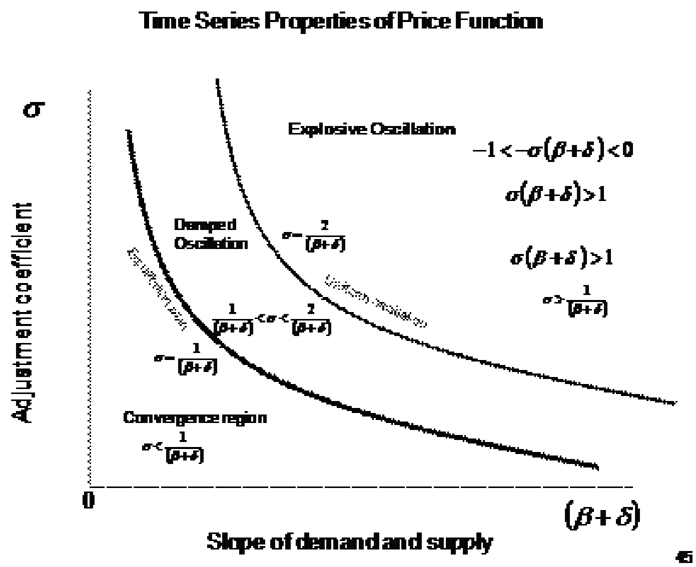
Inventory and Price Adjustment Model: Dynamic Properties

$$P_t = \left(P_0 - \frac{\alpha + \gamma}{\beta + \delta} \right) \left(1 - \sigma \left(\beta + \delta\right)\right)^t + \frac{(\alpha + \gamma)}{(\beta + \delta)} \tag{1945}$$

$b = \left(1 - \sigma \left(\beta + \delta\right)\right)$

1. $0 < b < 1$ convergent and non-oscillatory, $\sigma < \frac{1}{(\beta+\delta)}$.

2.$b=0$ solution is convergent to the steady state, $\sigma = \frac{1}{(\beta+\delta)}$
3.$-1<b<0$ gives oscillating but convergent path, $\frac{1}{(\beta+\delta)} < \sigma < \frac{2}{(\beta+\delta)}$
4. $b=-1$ case of regular oscillation, $\sigma = \frac{2}{(\beta+\delta)}$
5. $b<-1$ divergent oscillations, $\sigma > \frac{2}{(\beta+\delta)}$
Inventory and Price Adjustment Model: Dynamic Properties



18.2 First order differential equation

Difference equations are used to denote the time path of a variable when variables change continuously not discretely.

First order differential equation only involved differential term of order one.

$$\dot{y} + ay = b \ or \frac{\partial y}{\partial t} + ay = b \tag{1946}$$

Solution of a differential equation includes complementary and particular (steady state) parts

$$y_t = y_c + y_p \tag{1947}$$

For the steady state equilibrium $\dot{y} = 0$. This implies $y_p = \frac{b}{a}$
Solve the homogeneous system for the complementary solution:

$$\dot{y} + ay = 0 \tag{1948a}$$

$$\frac{\dot{y}}{y} = -a \tag{1949}$$

Solution of Differential Equation

Integrate both sides with respect to t

$$\int \frac{\dot{y}}{y} \partial t = \int -a \partial t \tag{1950}$$

$$ln\left(y_t\right) + c_{_1} = -at + c_{_2} \tag{1951}$$

Taking anti-log both sides

$$y_c = e^{-at} e^{c_2 - c_1} \tag{1952}$$

$$y_c = Ce^{-at} \tag{1953}$$

where $C = e^{c_2 - c_1}$
Complete solution

$$y_t = y_c + y_p = Ce^{-at} + \frac{b}{a} \tag{1954}$$

The time path of y_t converges if $a > 0$.
First Economic Example of the first order difference equation (IS-LM Model):
Consumption function :

$$C = a + BY - nR \tag{1955}$$

Let investment and government spending be as given at $I = \overline{I}$and $G = \overline{G}$

Goods markets does not balance automatically, it take time for adjustment as given by the following equation ($\alpha < 1$):

$$\frac{\partial y}{\partial t} = \alpha \left(a + by - nR + I + G - y\right) \tag{1956}$$

Money market is assumed to balance instantaneously

$$L = ky - hR \tag{1957}$$

$$L = \overline{M} \tag{1958}$$

Money market equilibrium implies

$$R = \frac{k}{h} y - \frac{1}{h} \overline{M} \tag{1959}$$

Putting the money market equilibrium in the goods market gives the economywide equilibrium process as:

$$\frac{\partial y}{\partial t} = \alpha \left(a + by - \left(\frac{nk}{h} y - \frac{n\overline{M}}{h}\right) + I + G - y\right) \tag{1960}$$

By rearrangement

$$\frac{\partial y}{\partial t}+\alpha\left(1-b+\frac{nk}{h}\right)y=\alpha\left(a+\frac{n\overline{M}}{h}+I+G\right) \tag{1961}$$

$$\frac{\partial y}{\partial t}+Ay=B \tag{1962}$$

where $A=\alpha\left(1-b+\frac{nk}{h}\right)$ and $B=\alpha\left(a+\frac{n\overline{M}}{h}+I+G\right)$The steady state equilibrium is given by $y_p=\frac{B}{A}=\frac{\alpha\left(a+\frac{n\overline{M}}{h}+I+G\right)}{\alpha\left(1-b+\frac{nk}{h}\right)}$ and the complementary path is given by

$$y_c=Ce^{-At}=Ce^{-\alpha\left(1-b+\frac{nk}{h}\right)t} \tag{1963}$$

Complete income path from solving the difference equation is given by combinations of these two:

$$y_t=Ce^{-At}+\frac{B}{A} \tag{1964}$$

Definite solution requires getting value of C using the initial conditions $y_{t=0}=y_0$ as $C=y_0-\frac{B}{A}$

$$\begin{aligned} y_t &= \left[y_0-\frac{B}{A}\right]e^{-At}+\frac{B}{A} \\ &= \left[y_0-\frac{\alpha\left(a+\frac{n\overline{M}}{M}+I+G\right)}{\alpha\left(1-b+\frac{nk}{h}\right)}\right]e^{-\alpha\left(1-b+\frac{nk}{h}\right)t}+\frac{\alpha\left(a+\frac{n\overline{M}}{h}+I+G\right)}{\alpha\left(1-b+\frac{nk}{h}\right)} \end{aligned} \tag{1965}$$

Convergence to the steady state requires that $A>0$. This implies $1-b+\frac{nk}{h}>0$ or $\frac{k}{h}>-\frac{1-b}{n}$ .The slope of the LM curve$\left(\frac{k}{h}\right)$ should be greater than the slope of the IS curve$\left(-\frac{1-b}{n}\right)$.

Consider a market price adjustment model where it takes time for demand and supply to adjust towards equilibrium. Starting from an initial point, does market prices converge to the long run equilibrium or not depends on the roots of the equations. These provide stability conditions for the system:

$$demand \quad Q^D=\alpha-\beta P \quad with \quad \alpha,\beta>0 \tag{1966}$$

$$Supply \quad Q^S=-\gamma+\delta P \quad with \quad \gamma,\delta>0 \tag{1967}$$

price adjustment process

$$\frac{\partial P}{\partial t}=k\left(\alpha-\beta P\ +\gamma-\delta P\right) \tag{1968}$$

by rearranging $\frac{\partial P}{\partial t}+k\left(\beta+\delta\right)P=k\left(\alpha\ +\gamma\right)$
The steady state equilibrium is $\overline{P}=\frac{\alpha\ +\gamma}{\beta+\delta}$
Homogeneous equation for complementary solution is given by:

$$\frac{\partial P}{\partial t}+k\left(\beta+\delta\right)P=0 \tag{1969}$$

$$\int \frac{\frac{\partial P}{\partial t}}{P} dt = -\int k(\beta+\delta)\, dt \tag{1970}$$

$$ln(P_t) + c_1 = -k(\beta+\delta)\, t + c_2 \tag{1971}$$

Taking anti-log both sides

$$P_t = e^{-k(\beta+\delta)t} e^{c_2 - c_1} \tag{1972}$$

$$P_t = Ce^{-k(\beta+\delta)t} \tag{1973}$$

Complete solution

$$P_t = P_c + P_p = Ce^{-k(\beta+\delta)t} + \frac{\alpha\ +\gamma}{\beta+\delta} \tag{1974}$$

18.3 Second order differential equation: market example

In addition to the structure of market above let the speculations in the demand side market determined by the first and second order conditions as following

$$demand \quad Q^D = \alpha - \beta P\ + mP' + nP'' \tag{1975}$$

$$Supply \quad Q^S = -\gamma + \delta P + uP' + wP'' \tag{1976}$$

for a while assume that $u = 0$ and $w = 0$

Let market find its equilibrium in each period $Q^D = Q^S$. This implies

$$\alpha - \beta P\ + mP' + nP''\ = -\gamma + \delta P \tag{1977}$$

$$nP'' + mP'\ - (\beta+\delta) P = -(\gamma+\alpha) \tag{1978}$$

The steady state equilibrium like before is : $P_p = \frac{\gamma+\alpha}{\beta+\delta}$

For complementary solution derive the homogenous equation

$$P'' + \frac{m}{n}P'\ - \left(\frac{\beta+\delta}{n}\right) P = 0 \tag{1979}$$

Let $P = Ae^{rt}$ sot that $P' = rAe^{rt}$ and $P'' = r^2Ae^{rt}$. and $r^2Ae^{rt} + \frac{m}{n}rAe^{rt} - \left(\frac{\beta+\delta}{n}\right) Ae^{rt} = 0$. The corresponding characteristic equations is:

$$r^2 + \frac{m}{n}r - \left(\frac{\beta+\delta}{n}\right) = 0 \tag{1980}$$

Roots of this equations are given by:

$$r_1, r_2 = \frac{-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta+\delta}{n}\right)}}{2} = \frac{1}{2}\left[-\frac{m}{n} \pm \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta+\delta}{n}\right)}\right] \tag{1981}$$

General solutions in the distinct real roots case when $\left[\left(\frac{m}{n}\right)^2 > 4\left(\frac{\beta+\delta}{n}\right)\right]$:

$$P_t = P_c + P_p = A_1 e^{r_1 t} + A_2 e^{r_2 t} + \frac{\alpha \ + \gamma}{\beta + \delta} \tag{1982}$$

It requires two initial conditions for definite solution

$$P_t = A_1 e^{\frac{1}{2}\left[-\frac{m}{n} - \sqrt{\left(\frac{m}{n}\right)^2 - 4\left(\frac{\beta+\delta}{n}\right)}\right]t} + A_2 e^{\frac{1}{2}\left[-\frac{m}{n} + \sqrt{\left(\frac{m}{n}\right)^2 + 4\left(\frac{\beta+\delta}{n}\right)}\right]t} + \frac{\alpha \ + \gamma}{\beta + \delta} \tag{1983}$$

In case of repeated root $\left(\frac{m}{n}\right)^2 = -4\left(\frac{\beta+\delta}{n}\right)$ there is only one root $r_1, r_2 = -\frac{m}{2n}$

$$P_t = P_c + P_p = A_3 e^{r_1 t} + A_4 t e^{r_2 t} + \frac{\alpha \ + \gamma}{\beta + \delta} \tag{1984}$$

for complex root case $\left[\left(\frac{m}{n}\right)^2 < -4\left(\frac{\beta+\delta}{n}\right)\right]$ the roots are divided between the real and imaginary parts as:

$$r_1, r_2 = h \pm vi \tag{1985}$$

where the real part in this case is $h = -\frac{m}{2n}$ and the $v = \left[-4\left(\frac{\beta+\delta}{n}\right) - \left(\frac{m}{n}\right)^2\right]$ and $i = \sqrt{-1}$.

Substituting real and imaginary parts and using the Euler equation and DeMoivre theorems:

$$P_t = P_c + P_p = e^{-\frac{m}{2n}t}\left[A_5 \cos(vt) + A_6 \operatorname{Si} n\,(vt)\right] + \frac{\alpha \ + \gamma}{\beta + \delta} \tag{1986}$$

Second order differential equation only involved differential term of order two. The procedure is similar to the second order difference equation. As before $y_t = y_c + y_p$ and $y_p = \frac{b}{a_2}$. For complementary solution $y = Ae^{rt}$ sot that $\dot{y} = rAe^{rt}$ and $\ddot{y} = r^2 Ae^{rt}$.

$$\ddot{y} + a_1 \dot{y} + a_2 y = b \tag{1987}$$

$$r^2 Ae^{rt} + a_1 r Ae^{rt} + a_2 Ae^{rt} = 0 \tag{1988}$$

$$r^2 + a_1 r + a_2 = 0 \tag{1989}$$

$$r_1, r_2 = \frac{(-a_1) \pm \sqrt{a_1^2 - 4a_2}}{2} \tag{1990}$$

There can be three cases in the solution of this equation depending on the value of the term under the square root

I. Distinct real root if $a_1^2 > 4a_2$

II. Repeated real root if $a_1^2 = 4a_2$

III. Complex real root $a_1^2 < 4a_2$ This requires use of the imaginary number, De Moivre theorem and trigonometry.

These cases is illustrated below by two examples:

Consider a market price adjustment model where it takes time for demand and supply to adjust towards equilibrium. Starting from an initial point, does market prices converge to the long run equilibrium or not depends on the roots of the equations. These provide stability conditions for the system:

Example of Complex Root Case: Example Preliminaries

Exponential forms and polar coordinates

$$R = \sqrt{h^2 + v^2} \tag{1991}$$

$$\sin n\theta = \frac{v}{R} \implies v = Rsin\theta \tag{1992}$$

$$\cos\theta = \frac{h}{R} \implies h = Rco\theta \tag{1993}$$

$$e^{i\theta} = \cos\theta + i\,\mathrm{Si}\,n\theta \qquad e^{-i\theta} = \cos\theta - i\,\mathrm{Si}\,n\theta \tag{1994}$$

$$h \pm vi = Rco\theta \pm Ri\sin\theta = R\left(co\theta \pm i\sin\theta\right) = \mathrm{Re}^{\pm i\theta} \tag{1995}$$

$\frac{\partial \sin\theta}{\partial\theta} = \cos\theta$; $\frac{\partial \cos\theta}{\partial\theta} = -\sin\theta$;

Thus the Cartesian coordinates of the complex numbers have been transformed to polar coordinates R and θ and also expressed as exponential form $\mathrm{Re}^{\pm i\theta}$.

Give the Cartesian form of the complex number $5\mathrm{e}^{3i\frac{\pi}{2}}$. Here R = 5, $\theta = 3\frac{\pi}{2}$

$R\left(co\theta \pm i\sin\theta\right) = 5\left(\cos 3\frac{\pi}{2} \pm i\sin 3\frac{\pi}{2}\right) = 5\left(\cos 0 \pm i\left(-1\right)\right) = -5i = h \pm vi$

By De Moivre's theorem

$(h + vi)^n = R^n e^{in\theta}$ *and* $(h - vi)^n = R^n e^{-in\theta}$

$$(h \pm vi)^n = R^n\left(\cos n\theta \pm i\sin n\theta\right) \tag{1996}$$

Solving a differential equation with complex roots

Table 23: Values of Trigonometric Ratios

	0^0	30^0	45^0	60^0	90^0	120^0	180^0	270^0	360^0
θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{3}{4}\pi$	π	$\frac{3}{2}\pi$	2π
sinθ	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1	$\frac{1}{\sqrt{2}}$	0	-1	0
cosθ	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0	-$\frac{1}{\sqrt{2}}$	-1	0	1

Example of Complex Root Case: Example

An Example

$$\ddot{y} + 2\dot{y} + 17y = 34 \tag{1997a}$$

Steady state

$$y_p = \frac{b}{a_2} = \frac{34}{17} = 2 \tag{1998}$$

This is a complex root case because $(a_1 = 2;\ a_2 = 17;\ b = 34)$

$a_1^2 - 4a_2 = 2^2 - 4 \times 17 = 4 - 68 = -64 < 0$

Use the formula explained above

$h \pm vi = Rco\theta \pm Ri\sin\theta = R\left(co\theta \pm i\sin\theta\right) = \mathrm{Re}^{\pm i\theta}$

$h = -\frac{1}{2}a_1 = -1 \quad v = \frac{1}{2}\sqrt{4a_2 - a_1^2} = \frac{1}{2}\sqrt{4\,(17) - 2^2} = \frac{1}{2}\sqrt{64} = \frac{1}{2}\,(8) = 4$
In case of the complex root

$$\begin{aligned} y_c &= e^{ht}\left(A_1 e^{vit} + A_2 e^{-vit}\right) && (1999) \\ &= e^{ht}\left[A_1\left(\cos vt + i\sin vt\right) + A_2\left(\cos vt - i\sin vt\right)\right] && (2000) \end{aligned}$$

For this problem complementary solution

$$\begin{aligned} y_c &= e^{ht}\left(A_1 e^{4it} + A_2 e^{-4it}\right) && (2001) \\ &= e^{-t}\left[A_1\left(\cos 4t + i\sin 4t\right) + A_2\left(\cos 4t - i\sin 4t\right)\right] && (2002) \end{aligned}$$

$$y_t = y_c + y_p = e^{-t}\left[A_1\left(\cos 4t + i\sin 4t\right) + A_2\left(\cos 4t - i\sin 4t\right)\right] + 2 \qquad (2003)$$

$$y_t = e^{-t}\left[\left(A_1 + A_2\right)\cos 4t + \left(A_1 - A_2\right) i\sin 4t\right] + 2 \qquad (2004)$$

$$y_t = e^{-t}\left[A_5\cos 4t + A_6\sin 4t\right] + 2 \qquad (2005)$$

where $A_5 = (A_1 + A_2) \qquad A_6 = (A_1 - A_2)\,i$
Use two initial conditions to definitize the values of A_5 and A_6.
$y_0 = 3$ and $\dot{y} = 11$.
When $t = 0$

$$y_0 = 3 = e^{-t}\left[A_5\cos 4t + A_6\sin 4t\right] + 2 = \left[A_5\cos 0 + A_6\sin 0\right] + 2 = A_5 + 2 \qquad (2006)$$

Thus $\mathrm{A}_5 = 1$
take the first derivative of with respect to time

$$\begin{aligned} \dot{y} &= \frac{\partial y}{\partial t}\left\{e^{-t}\left[A_5\cos 4t + A_6\sin 4t\right] + 2\right\} && (2007) \\ \dot{y} &= -e^{-t}\left[A_5\cos 4t + A_6\sin 4t\right] + e^{-t}\left[-4A_5\sin 4t + 4A_6\cos 4t\right] && (2008) \end{aligned}$$

Evaluated when $t = 0$
$\dot{y} = -e^{-t}\left[A_5\cos 0 + A_6\sin 0\right] + e^{-t}\left[-4A_5\sin 0 + 4A_6\cos 0\right]$

$$11 = -\left(A_5 + 0\right) + \left[0 + 4A_6\right] \qquad (2009)$$

$\mathrm{A}_6 = 3$
Thus the complete solution of this equation is:

$$y_t = e^{-t}\left[\cos 4t + 3\sin 4t\right] + 2 \qquad (2010)$$

The first trigonometric function gives the cycle and second part is the steady state.
Numerical example 1 for SODE

$$demand \quad Q^D = 42 - 4P \ + 4P' + P'' \tag{2011}$$

$$Supply \quad Q^S = -6 + 8P \tag{2012}$$

Initial conditions $P_0 = 6$ and $P'(t = 0) = 4$.
Let market find its equilibrium in each period $Q^D = Q^S$. This implies

$$42 - 4P \ + 4P' + P'' \ = -6 + 8P \tag{2013}$$

The steady state equilibrium like before is : $P_p = \frac{46}{12} = 4$
For homogenous solution rearrange $P'' - 4P' \ - 4P + 42 = -6 + 8P$ to

$$P'' - 4P' \ - 8P = 0 \tag{2014}$$

Numerical example 1 for SODE Corresponding quadratic equation is given by

$$r_1, r_2 = \frac{-(-4) \pm \sqrt{(-4)^2 - 4.1.(-12)}}{2} = \frac{4 \pm \sqrt{16 + 46}}{2} = 6, -2 \tag{2015}$$

$$P_t = P_c + P_p = A_1 e^{r_1 t} + A_2 e^{r_2 t} + 4 = A_1 e^{6t} + A_2 e^{-2t} + 4 \tag{2016}$$

Use two initial conditions for the complete solution

$$P_0 = 6 = A_1 e^{6.0} + A_2 e^{-2.0} + 4 = A_1 + A_2 + 4 \tag{2017}$$

$$P' = 4 = 6A_1 e^{6.0} - 2A_2 e^{-2.0} = 6A_1 - 2A_2 \tag{2018}$$

Solving these equations $A1 = 1$ and $A2 = 1$.

$$P_t = A_1 e^{r_1 t} + A_2 e^{r_2 t} + 4 = e^{6t} + e^{-2t} + 4 \tag{2019}$$

This path is dynamically unstable because of $r_1 = 6$. This gives divergent Oscillations.

Numerical example 2 for SODE

$$demand \quad Q^D = 40 - 2P \ - 2P' - P'' \tag{2020}$$

$$Supply \quad Q^S = -5 + 3P \tag{2021}$$

Initial conditions $P_0 = 12$ and $P'(t = 0) = 1$.
Let market find its equilibrium in each period $Q^D = Q^S$. This implies

$$40 - 2P \ - 2P' - P'' \ = -5 + 3P \tag{2022}$$

The steady state equilibrium like before is : $P_p = \frac{45}{5} = 9$
For homogenous solution rearrange $40 - 2P \ - 2P' - P'' \ = -5 + 3P$ to$P'' + 2P' + 5P \ = 45$

$$r_1, r_2 = \frac{-2 \pm \sqrt{2^2 - 4.1.5}}{2} = \frac{-2 \pm \sqrt{4-20}}{2} = \frac{1}{2}(-2 \pm 4i) = -1 \pm 2i \tag{2023}$$

This is complex root case with $h + vi = -1 \pm 2i$ where $h = -1$ and $v = 2$
The general solution of this model is

$$P_t = P_c + P_p = e^{-t}\left[A_5 \cos(2t) + A_6 \operatorname{Si} n\,(2t)\right] + 9 \tag{2024}$$

Using the initial conditions

$$P_0 = 12 = e^{-0}\left[A_5 \cos(0) + A_6 \operatorname{Si} n\,(0)\right] + 9 = A_5(1) + A_6.0 + 9 = A_5 + 9 \tag{2025}$$

$$P'_t = -e^{-t}\left[A_5 \cos(2t) + A_6 \operatorname{Si} n\,(2t)\right] + e^{-t}\left[-2A_5 \sin(2t) + 2A_6 Cos\,(2t)\right] \tag{2026}$$

$$\begin{aligned} P'_{t=0} &= 1 \\ &= -e^{-0}\left[A_5 \cos(2.0) + A_6 \operatorname{Si} n\,(2.0)\right] \\ &\quad + e^{-0}\left[-2A_5 \sin(2.0) + 2A_6 Cos\,(2.0)\right] \\ &= A_5 + 0 + 0 + 2A_6 \end{aligned} \tag{2027}$$

Solving $A_5 + 9 = 12$ and $A_5 + 2A_6 = 1$ we get $A_5 = 3$ and $A_6 = 2$. Thus the definite solution path of the system is

$$P_t = e^{-t}\left[3\cos(2t) + 2 \operatorname{Si} n\,(2t)\right] + 9 \tag{2028}$$

P_t fluctuates in each period of $\frac{2\pi}{v} = \pi = 3.1452$. when t increases 3.1452 the P_t completes one cycle.

This cycle is damped because of the multiplicative term e^{-t}.

That means this path P_t starts at 12 and gradually converges to 9 in a cyclical fashion.

18.3.1 Generic Differential Equations: Routh Theorem

In a higher order differential equation Routh theorem is applied to find whether time path converges to long run equilibrium:

Take a polynomial of the form
$a_0 r^n + a_1 r^{n-1} + a_2 r^{n-2} + ... + a_{n-1} r + a_n = 0$
the real parts of all the roots of nth degree polynomial are negative when first n sequence of determinants are positive. Therefore above equation is convergent.

Generic Differential Equations

Routh Matrix is formed by letting odd coefficients head a row and successively reducing the subscripts and writing zero for negative coefficients (Samuelson (1947) Foundations of Economic Analysis).

$$|a_1|; \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix} \begin{vmatrix} a_1 & a_3 & a_5 & a_7 \\ a_0 & a_2 & a_4 & a_6 \\ 0 & a_1 & a_3 & a_5 \\ 0 & a_0 & a_2 & a_4 \end{vmatrix}$$

Generic Differential Equations

Numerical example $y^4(t)+6y^{'''}(t)+14y^{''}(t)+16y^{'}(t)+8y=24$

$a_0=1; a_1=6; a_2=14; a_3=16; a_4=8; a_5=0; a_6=0;$

$\Delta_0=|a_1|=|6|>0; \Delta_1=\begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}=\begin{vmatrix} 6 & 16 \\ 1 & 14 \end{vmatrix}=84-16=68>0; \begin{vmatrix} 6 & 16 & 0 \\ 1 & 14 & 8 \\ 0 & 6 & 16 \end{vmatrix}=800>0;$

$\begin{vmatrix} 6 & 16 & 0 & 0 \\ 1 & 14 & 8 & 0 \\ 0 & 6 & 16 & 0 \\ 0 & 1 & 14 & 8 \end{vmatrix}=6400>0$

The first n sequence of determinants are positive, the real parts of all the roots of nth degree polynomial are negative . Therefore the time path of $y(t)$ in above equation is convergent.

Application of Difference Equation ARCH(q) Process:

$$y_t=\mu+u_t \tag{2029}$$

$$u_t=h_t^{\frac{1}{2}}\varepsilon_t \quad \epsilon_t\sim N\left(0,\sigma^2\right) \tag{2030}$$

$$h_t=\alpha_0+\sum_{j=1}^{q}\alpha_j u_{t-j}^2 \tag{2031}$$

Log likelihood of ARCH function for observation t is

$$lnL(\mu,\alpha_1,\alpha_2....\alpha_n)=c-\frac{1}{2}\ln h_t-\frac{1}{2}\frac{(y_t-\mu)^2}{h_t} \tag{2032}$$

$$lnL(\mu,\alpha_1,\alpha_2....\alpha_n)=c-\frac{T}{2}\sum_{t=1}^{T}\ln h_t-\frac{T}{2}\sum_{t=1}^{T}\frac{(y_t-\mu)^2}{h_t} \tag{2033}$$

this leads to the estimation of$c=-\frac{T}{2}\ln(2\pi)$, $\mu,\alpha_1,\alpha_2....\alpha$ and $\widehat{u}_t$,$\widehat{u}_{t-1,}$ $\widehat{u}_{t-1}$ helps to estimate $\widehat{\sigma}^2$.

GARCH(p,q)

$$y_t=x_t^{'}\xi+u_t \tag{2034}$$

$$u_t=h_t^{\frac{1}{2}}\varepsilon_t \quad \epsilon_t\sim N\left(0,\sigma^2\right) \tag{2035}$$

$$h_t=\alpha_0+\sum_{j=1}^{q}\alpha_j u_{t-j}^2+\sum_{j=1}^{p}\beta_j h_{t-1} \tag{2036}$$

Using the lag operator

$$h_t=\frac{\alpha_0}{1-\beta(L)}+\frac{\alpha(L)}{1-\beta(L)_j}u_t^2 \tag{2037}$$

ARCH(q) is equivalent to GARCH(0,q)
PcGive Volatility package computes ARCH and GARCH
Heteroscedasticy adjusted mean square error

$$HMSE = \frac{1}{T}\sum\left(\frac{\widehat{u}_t^2}{\widehat{h}_t} - 1\right)^2 \tag{2038}$$

Mean of the conditional variance

$$\overline{h}_t = \frac{1}{T}\sum_{t=1}^{T}\widehat{h}_t \ ; \ \ var\left(h_t\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{h}_t - \overline{h}_t\right) \tag{2039}$$

AIC criteria

$$AIC \times T = -2\widehat{l} + 2s; \ AIC = -\frac{2\widehat{l}}{T} + \frac{2s}{T} \tag{2040}$$

Exponential and integrated GARCH are varieties of other GARCH models.

References

[1] Chiang A.C. (1984) Fundamental Methods of Mathematical Economics, 3rd edition, McGraw Hill.

[2] Enders W. (2010) Applied Econometric Time Series, 3rd ed., Willey.

[3] Greene W. (2008) Econometric Analysis, Fifth Edition, Prentice Hall.

[4] Hoy M., J. Livernois, C McKenna, R. Rees and T Stengos (2001) Mathematics for Economics, MIT Press

18.3.2 Tutorial 1: Optimisation and Matrix

Q1. Consider y as a function of x_1, x_2 and x_3 as given in the following equation:

$$y = -5x_1^2 + 10x_1 + x_1x_2 - 2x_2^2 + 4x_2 + 2x_2x_3 - 4x_3^2 \tag{2041}$$

a. Find the optimal values of , and using the first order conditions for unconstrained maximisation. Use matrix approach in your solution.

b. Determine whether the above solutions correspond to the minimum or the maximum point using positive or negative definite concepts of the Hessian determinants.

Q2. Consider coefficients of a market model given by a matrix $A = \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}$

a) What are the eigen values of this maxtrix?

b) What are associated eigen vectors?

c) Prove that eigen vectors are orthogonal, $\left(V_1\right)^{'}\left(V_2\right) = 0$

d) Prove $\left(V_1V_2\right)^{'}\left(V_1V_2\right) = \left(V_1V_2\right)\left(V_1V_2\right)^{'} = I$

Q3. Write short notes in any three of the following.

- **a.** Order of integration and unit root test.
- **b.** Engle-Granger Representation theorem.
- **c.** Johansen test for cointegrating vector.
- **d.** Simultanety bias.
- **e.** Autocorrelation.

18.3.3 Higher Order Difference Equations: Schurr Theorem

Checking convergence of a difference equation (Schur determinants approach)

$$Y_{t+2} + \frac{1}{6}Y_{t-1} - \frac{1}{6}Y_t = 2 \tag{2042}$$

This is a second order difference equation
$a_0 = 1$; $a_1 = \frac{1}{6}$; $a_2 = -\frac{1}{6}$
; $\Delta_1 = \begin{vmatrix} a_0 & a_2 \\ a_2 & a_0 \end{vmatrix} > 0$; $\Delta_1 = \begin{vmatrix} 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 \end{vmatrix} = \frac{35}{36} > 0$

$$\Delta_2 = \begin{vmatrix} a_0 & 0 & a_2 & a_1 \\ a_1 & a_0 & 0 & a_2 \\ a_2 & 0 & a_0 & a_1 \\ a_1 & a_2 & 0 & a_0 \end{vmatrix} = \begin{vmatrix} 1 & -\frac{1}{6} & 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 & -\frac{1}{6} & 1 \\ 1 & -\frac{1}{6} & 1 & -\frac{1}{6} \\ -\frac{1}{6} & 1 & -\frac{1}{6} & 1 \end{vmatrix} = 0907407 > 0$$

Divide the matrix in four parts:

A	B
C	D

Start with a_0 in diagonal at the upper left matrix (A), put zeros above the diagonal and successively higher subscripts down the column (A)

Matrix at the southeast corner (D) is the transpose of the northwest corner (A');

Put a_n in the diagonal of the south west corner (C) and zeros above the diagonal and successively smaller subscripts down the column of (C)

The matrix at northeast corner (B) is transpose of matrix at the southwest corner (C)

Roots of the polynomial are less than unity when Schur determinants are positive. Therefore above difference equation gives a convergent path.

Routh theorem used for differential equations as explained above.

18.3.4 Samuelsonian Multiplier Accelerator Model (Second Order Difference Equation)

Macro balance

$$Y_t = C_t + I_t + G_0 \tag{2043}$$

Consumption function

$$C_t = \gamma Y_{t-1}; \qquad 0 < \gamma < 1 \tag{2044}$$

Investment

$$I_t = \alpha \left(C_t - C_{t-1} \right); \qquad \alpha > 1 \tag{2045}$$

Equilibrium (putting C_t and I_t in Y_t): second order difference equation

$$Y_t = \gamma \left(1 + \alpha \right) Y_{t-1} - \gamma \alpha Y_{t-2} + G_0 \tag{2046}$$

Samuelsonian Multiplier Accelerator Model (Second Order Difference Equation)
Steady state output

$$\overline{Y} = \frac{G_0}{1 - \gamma \left(1 + \alpha \right) + \gamma \alpha} = \frac{G_0}{1 - \gamma} \tag{2047}$$

Homogenous second order difference equation

$$Y_t - \gamma \left(1 + \alpha \right) Y_{t-1} + \gamma \alpha Y_{t-2} = 0 \tag{2048}$$

Consumption function

$$C_t = \gamma Y_{t-1}; \qquad 0 < \gamma < 1 \tag{2049}$$

Investment

$$I_t = \alpha \left(C_t - C_{t-1} \right); \qquad \alpha > 1 \tag{2050}$$

Equilibrium (putting C_t and I_t in Y_t)

$$Y_t = \gamma \left(1 + \alpha \right) Y_{t-1} - \gamma \alpha Y_{t-2} + G_0 \tag{2051}$$

Solution of the Samuelsonian Multiplier Accelerator Model
Steady state output

$$\overline{Y} = \frac{G_0}{1 - \gamma \left(1 + \alpha \right) + \gamma \alpha} = \frac{G_0}{1 - \gamma} \tag{2052}$$

Transitional dynamics (replace $Y_t = Ab^t$ in homogenous equation).

$$Y_t - \gamma \left(1 + \alpha \right) Y_{t-1} + \gamma \alpha Y_{t-2} = 0 \tag{2053}$$

$$Ab^t - \gamma \left(1 + \alpha \right) Ab^{t-1} + \gamma \alpha Ab^{t-2} = 0 \tag{2054}$$

$$b^2 - \gamma \left(1 + \alpha \right) b + \gamma \alpha = 0 \tag{2055}$$

Cycle depends on roots of the quadratic equation

$$b_1, b_2 = \frac{\gamma (1+\alpha) \pm \sqrt{\gamma^2 (1+\alpha)^2 - 4\gamma\alpha}}{2} \tag{2056}$$

Three Cases in Samuelsonian Multiplier Accelerator Model
Distinct real root case (no cycle)

$$\gamma^2 (1+\alpha)^2 > 4\gamma\alpha \tag{2057}$$

Repeated real root case (no cycle)

$$\gamma^2 (1+\alpha)^2 = 4\gamma\alpha \tag{2058}$$

Complex root case (cycle)

$$\gamma^2 (1+\alpha)^2 < 4\gamma\alpha \tag{2059}$$

Complete solution

$$Y_t = A_1 b_1^t + A_2 b_2^t + \overline{Y} \tag{2060}$$

$Y_t = A_1 R^t (\cos\theta \cdot t + i \cdot \sin\theta \cdot t) + A_2 R^t (\cos\theta \cdot t - i \cdot \sin\theta \cdot t) + \overline{Y}$
Multiplier Accelerator Model
Two roots of a characteristic equation are related as:

$$b_1 + b_2 = \gamma (1+\alpha) \tag{2061}$$

$$b_1 b_2 = \gamma\alpha \tag{2062}$$

Using these consider how the values of characteristic root compare to the values of α and γ.

$$(1-b_1)(1-b_2) = 1 - (b_1 + b_2) + b_1 b_2 \tag{2063}$$

$$(1-b_1)(1-b_2) = 1 - \gamma (1+\alpha) + \gamma\alpha = 1 - \gamma \tag{2064}$$

Condition $0 < (1-b_1)(1-b_2) < 1$ is necessary to remain consistent with the potential value of the marginal propensity to consume, $0 < \gamma < 1$.

Two roots of a characteristic equation are related as:

$$b_1 + b_2 = \gamma (1+\alpha) \tag{2065}$$

$$b_1 b_2 = \gamma\alpha \tag{2066}$$

Using these consider how the values of characteristic root compare to the values of α and γ.

$$(1-b_1)(1-b_2) = 1 - (b_1 + b_2) + b_1 b_2 \tag{2067}$$

$$(1-b_1)(1-b_2) = 1 - \gamma (1+\alpha) + \gamma\alpha = 1 - \gamma \tag{2068}$$

Condition $0 < (1-b_1)(1-b_2) < 1$ is necessary to remain consistent with the potential value of the marginal propensity to consume, $0 < \gamma < 1$.

There are five possible configurations of and their implications on multiplier and acceleration terms and are as following:

1. $0 < b_2 < b_1 < 1 \Longrightarrow 0 < \gamma < 1$ and $\gamma\alpha < \mathbf{1}$ Convergent

2.$0 < b_2 < b_1 = 1 \Longrightarrow \gamma = 1$ MPC should be less than 1

3. $0 < b_2 < 1 < b_1 \Longrightarrow \gamma > 1$ MPC should be less than 1

4. $0 < b_2 = 1 < b_1 \Longrightarrow \gamma > 1$ MPC should be less than 1

5.$1 = b_2 < b_1 \Longrightarrow 0 < \gamma < 1$ and $\gamma\alpha > \mathbf{1}$ Divergent

Multiplier Accelerator Model: Real Distinct Roots

The convergence of the system depends on term $\gamma\alpha$. System is convergent $\gamma\alpha < \mathbf{1}$, has steady if $\gamma\alpha = \mathbf{1}$or divergent $\gamma\alpha > \mathbf{1}$.

The degree of fluctuation depends on $\frac{4\alpha}{(1+\alpha)^2}$ relative to γ. The system is explosive with no oscillation if $\gamma > \frac{4\alpha}{(1+\alpha)^2}$;

it is recurrent if $\gamma = \frac{4\alpha}{(1+\alpha)^2}$and has damped oscillations if $\gamma < \frac{4\alpha}{(1+\alpha)^2}$.

This last case requires solving the model using a complex root.

Multiplier Accelerator Model: Repeated Roots

$\gamma^2 (1+\alpha)^2 = 4\gamma\alpha$

When there is only one solution for both $b_1$and b_2.

There can be three cases in this situation.

$$b_1, b_2 = \frac{\gamma(1+\alpha) \pm \sqrt{\gamma^2 (1+\alpha)^2 - 4\gamma\alpha}}{2} = \frac{\gamma(1+\alpha)}{2} \tag{2069}$$

$1 = b_2 < b_1 \Longrightarrow 0 < \gamma < 1$ and $\gamma\alpha > \mathbf{1}$

1. $0 < b_1 < 1 \Longrightarrow < \gamma < 1$ and $\gamma\alpha < \mathbf{1}$ convergence no oscillation.$0 < (1-b_1)(1-b_2) < 1$

2. $0 < b_1 = 1 \Longrightarrow \gamma = 1$ violates condition $0 < (1-b_1)(1-b_2) < 1$

3. $0 < b_1 = 1 \Longrightarrow < \gamma < 1$ $\gamma\alpha > \mathbf{1}$ divergent no oscillation

Multiplier Accelerator Model: Complex Root

$\gamma^2 (1+\alpha)^2 < 4\gamma\alpha$

Need to consider the algebra for the imaginary number and some trigonometric functions in this case. Using Pythagorean in an imaginary axis is used to derive the roots of the characteristic equation.

$$b_1, b_2 = (h \pm v \cdot i) = \frac{\gamma(1+\alpha)}{2} \pm i\sqrt{\frac{4\gamma\alpha - \gamma^2 (1+\alpha)^2}{2}} \tag{2070}$$

$$Y_t = A_1 b_1^t + A_2 b_2^t = A_1 (h + v \cdot i)^t + A_2 (h - v \cdot i)^t \tag{2071}$$

Using DeMoivre's theorem

$$(h \pm v \cdot i) = R^t (\cos\theta \cdot t \pm i \sin\theta \cdot t) \quad \text{for} \quad R^t > 0. \tag{2072}$$

Multiplier Accelerator Model: Complex Root Imaginary axis (Pithagorus Theorem)

$$R = \sqrt{h^2 + v^2} = \alpha\gamma \tag{2073}$$

Coefficient of Yt-2

$$Y_t = A_1 R^t (\cos\theta \cdot t + i \sin\theta \cdot t) + A_2 R^t (\cos\theta \cdot t - i \sin\theta \cdot t) \tag{2074}$$

$$Y_t = A_1 R^t \left(\cos\frac{\pi}{2} \cdot t + i \sin\frac{\pi}{2} \cdot t\right) + A_2 R^t \left(\cos\frac{\pi}{2} \cdot t - i \sin\frac{\pi}{2} \cdot t\right) \tag{2075}$$

Three possibilities:

i) $R^t > 1$; $\alpha\gamma > 1$ ii) $R^t = 1 \quad \alpha\gamma = 1$ and ii) $R^t < 1$ $\alpha\gamma < 1$ Only the $\alpha\gamma < 1$ case is convergent other two cases are divergent.

18.4 Review of matrix algebra

$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}; \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}; \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix};$

Addition:

$$A + B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} \tag{2076}$$

Subtraction:

$$A - B = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} - b_{11} & a_{12} - b_{12} \\ a_{21} - b_{21} & a_{22} - b_{22} \end{bmatrix} \tag{2077}$$

Multiplication:

$$AB = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} \tag{2078}$$

18.4.1 Determinant and Transpose of Matrices

Determinant of A

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = (a_{11}a_{22} - a_{21}a_{12})\,; \tag{2079}$$

Determinant of B $\quad |B| = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = (b_{11}b_{22} - b_{21}b_{12})$

Determinant of C $\quad |C| = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} = (c_{11}c_{22} - c_{21}c_{12})$

Transposes of A, B and C

$$A' = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}; \; B' = \begin{bmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{bmatrix}; \; C' = \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \tag{2080}$$

Singular matrix $|D| = 0$. non-singular matrix $|D| \neq 0$.

18.4.2 Inverse of A

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{|A|} adj\,(A) \tag{2081}$$

$$adj\,(A) = C' \tag{2082}$$

For C cofactor matrix. For this cross the row and column corresponding to an element and multiply by $(-1)^{i+j}$

$$C = \begin{bmatrix} |a_{22}| & -|a_{21}| \\ -|a_{12}| & |a_{11}| \end{bmatrix} = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix} \tag{2083}$$

$$C' = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}' = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{2084}$$

Inverse of A

$$A^{-1} = \frac{1}{(a_{11}a_{22} - a_{21}a_{12})} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} = \begin{bmatrix} \frac{a_{22}}{(a_{11}a_{22}-a_{21}a_{12})} & -\frac{a_{12}}{(a_{11}a_{22}-a_{21}a_{12})} \\ -\frac{a_{21}}{(a_{11}a_{22}-a_{21}a_{12})} & \frac{a_{11}}{(a_{11}a_{22}-a_{21}a_{12})} \end{bmatrix} \tag{2085}$$

1) Find B^{-1}.
2) Some examples (a) addition and subtraction

18.4.3 Exercise on matrix manipulations

An electronic store has branches both in Hull and York and sells computers and TV before and after the Christmas. Quantities and prices are as given below.

Table 24: Hypothetical Data on Quantities and Prices

	Hull				York			
	Computer		TV		Computer		TV	
	Before	After	Before	After	Before	After	Before	After
Quantities (Y_i)	300	500	600	400	300	500	600	800
Prices (X_i)	500	400	100	60	525	400	120	80

Represent quantities and prices in the matrix form.

1. (a) Total quantities and prices sold in both markets before and after, i.e. (Q_B) and (Q_A) and (P_B) and (P_A).
   (b) Difference in sales in these two places $(Q_B - Q_A)$.
   (c) Total sales revenue in both places before and after the Christmas ($R_B = Q_B P_B$, $R_A = Q_A P_A$).
   (d) If total revenue (R_B, R_A) and quantities (Q_B, Q_A) are known, show formula to find prices (P_B, P_A).

3) Portfolio of A, B, C, and D companies is $P = \begin{bmatrix} 200 & 300 & -1100 & 600 \end{bmatrix}$ and their prices in good and bad economic states $S' = \begin{bmatrix} G & 1.3 & 1.2 & 1.0 & 1.5 \\ B & 1.5 & 0.83 & 0.95 & 1.2 \end{bmatrix}$ are respectively. Find the expected values of above portfolio in good and bad states (PS).

18.4.4 Application: Problems with multiple markets

Market 1:

$$X_1^d = 10 - 2p_1 + p_2 \tag{2086}$$

$$X_1^S = -2 + 3p_1 \tag{2087}$$

- Market 2:

$$X_2^d = 15 + p_1 - p_2 \tag{2088}$$

$$X_2^S = -1 + 2p_2 \tag{2089}$$

$X_1^d = X_1^S$ implies $10 - 2p_1 + p_2 = -2 + 3p_1$
$X_1^d = X_1^S$ implies $15 + p_1 - p_2 = -1 + 2p_2$

$$\begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 16 \end{bmatrix} \tag{2090}$$

18.4.5 Application of Matrix in Solving Equations

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 12 \\ 16 \end{bmatrix} \tag{2091}$$

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix} = (5 \times 3 - (-1)(-1)) = 15 - 1 = 14;$$

$$C' = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix}' = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}$$

$$\begin{aligned} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} &= \frac{1}{14} \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 12 \\ 16 \end{bmatrix} \\ &= \frac{1}{14} \begin{pmatrix} (3 \times 12) + (1 \times 16) \\ (1 \times 12) + (5 \times 16) \end{pmatrix} = \begin{pmatrix} \frac{52}{14} \\ \frac{92}{14} \end{pmatrix} = \begin{pmatrix} \frac{26}{7} \\ \frac{46}{7} \end{pmatrix} \end{aligned} \tag{2092}$$

18.5 Application of Matrix in Solving Equations

18.5.1 Cramer's Rule

$$p1 = \frac{\begin{vmatrix} 12 & -1 \\ 16 & 3 \end{vmatrix}}{\begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix}} = \frac{36 + 16}{15 - 1} = \frac{26}{7}; \quad p2 = \frac{\begin{vmatrix} 5 & 12 \\ -1 & 16 \end{vmatrix}}{\begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix}} = \frac{80 + 12}{15 - 1} = \frac{46}{7} \tag{2093}$$

Market 1:

$$LHS = 10 - 2p_1 + p_2 = 10 - 2\left(\frac{26}{7}\right) + \left(\frac{46}{7}\right) = \frac{64}{7} = -2 + 3p_1 = \frac{64}{7} = RHS \tag{2094}$$

Market 2:

$$LHS = 15 + p_1 - p_2 = 15 + \frac{26}{7} - \frac{46}{7} = \frac{85}{7} = -1 + 2p_2 = \frac{85}{7} = RHS \tag{2095}$$

QED.
Extension to N-markets is obvious; a confidence for solving large models.

18.5.2 Spectral Decomposition of a Matrix

$$|A - \lambda I| = \begin{vmatrix} 5 & -1 \\ -1 & 3 \end{vmatrix} - \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} = \begin{vmatrix} 5-\lambda & -1 \\ -1 & 3-\lambda \end{vmatrix} = 0 \tag{2096}$$

λ is Eigen value

$$(5-\lambda)(3-\lambda) - 1 = 0 \tag{2097}$$

$15 - 5\lambda - 3\lambda + \lambda^2 - 1 = 0$ or

$$\lambda^2 - 8\lambda + 14 = 0 \tag{2098}$$

Eigen values

$$\lambda_1, \lambda_2 = \frac{8 \pm \sqrt{8^2 - 4 \times 14}}{2} = \frac{8 \pm \sqrt{8}}{2} = \frac{8 \pm 2.83}{2} == 5.4, 2.6 \tag{2099}$$

$$\begin{bmatrix} 5-\lambda & -1 \\ -1 & 3-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2100}$$

for $\lambda_1 = 5.4$

$$\begin{bmatrix} 5-5.4 & -1 \\ -1 & 3-5.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2101}$$

$$\begin{bmatrix} -0.4 & -1 \\ -1 & -2.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2102}$$

$x_2 = -0.4x_1$
Normalisation

$$x_1^2 + x_2^2 = 1 \quad ; \quad x_1^2 + (-0.4x_1)^2 = 1 \tag{2103}$$

$$1.16x_1^2 = 1; x_1^2 = \frac{1}{1.16}; x_1 = \sqrt{0.862} = 0.928 \tag{2104}$$

$$x_2 = -0.4x_1 = -0.4\,(0.928) = -0.371 \tag{2105}$$

18.5.3 Eigenvector 1

$$V_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0.928 \\ -0.371 \end{pmatrix} \tag{2106}$$

$\lambda_2 = 2.6$

$$\begin{bmatrix} 5-2.6 & -1 \\ -1 & 3-2.6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2107}$$

$$\begin{bmatrix} 2.4 & -1 \\ -1 & 0.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2108}$$

$x_2 = 2.4x_1$

$$x_1^2 + x_2^2 = 1; \qquad x_1^2 + (2.4x_1)^2 = 1 \tag{2109}$$

$$6.76x_1^2 = 1; x_1^2 = \frac{1}{6.76}; x_1 = \sqrt{0.129} = 0.373 \tag{2110}$$

$$x_2 = 2.4 \;\; x_1 = 2.4\,(0.373) = 0.895 \tag{2111}$$

Eigenvector 2

$$V_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0.373 \\ 0.895 \end{pmatrix} \tag{2112}$$

Orthogonality (Required for GLS)

$$(V_1)'\,(V_2) = 0 \tag{2113}$$

$$V_1 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0.928 \\ -0.371 \end{pmatrix} \tag{2114}$$

$$V_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{pmatrix} 0.373 \\ 0.895 \end{pmatrix} \tag{2115}$$

$$\begin{bmatrix} 0.928 & -0.371 \end{bmatrix} \begin{bmatrix} 0.373 \\ 0.895 \end{bmatrix} = 0.346 - 0.332 \approx 0 \tag{2116}$$

18.5.4 Orthogonality (Required for GLS)

$$(V_1)'\,(V_2) = 0 \tag{2117}$$

$$\begin{bmatrix} 0.928 & -0.371 \end{bmatrix} \begin{bmatrix} 0.373 \\ 0.895 \end{bmatrix} = 0.346 - 0.332 \approx 0 \tag{2118}$$

$$(V_1V_2)'\,(V_1V_2) = (V_1V_2)\,(V_1V_2)' = I \tag{2119}$$

$$\begin{bmatrix} 0.928 & 0.373 \\ -0.371 & 0.895 \end{bmatrix} \begin{bmatrix} 0.928 & 0.373 \\ -0.371 & 0.895 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2120}$$

18.5.5 Diagonalisation, Trace of Matrix

Inverse of an orthogonal matrix equals its transpose $Q = \left(V_1^{'} V_2\right)$

$$Q^{-1} = Q' \tag{2121}$$

$$Q' A Q = \Lambda \tag{2122}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & .. & 0 \\ 0 & \lambda_2 & .. & 0 \\ : & : & : & : \\ 0 & 0 & .. & \lambda_n \end{bmatrix} \tag{2123}$$

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} \tag{2124}$$

$$|A| = \lambda_1 \lambda_2 .... \lambda_n \tag{2125}$$

18.5.6 Quadratic forms, Positive and Negative Definite Matrices

quadratic form

$$q\left(x\right) = \left(x_1 x_2\right) \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2126}$$

Positive definite matrix (matrix with all positive eigen values)

$$q\left(x\right) = x^{'} A x > 0 \tag{2127}$$

Positive semi-definite matrix

$$q\left(x\right) = x^{'} A x \geqslant 0 \tag{2128}$$

Negative definite matrix (matrix with all negative eigen values)

$$q\left(x\right) = x^{'} A x < 0 \tag{2129}$$

Negative semi-definite matrix

$$q\left(x\right) = x^{'} A x \leq 0 \tag{2130}$$

18.5.7 Generalised Least Square

Take a regression

$$Y = X\beta + e \tag{2131}$$

Assumption of homoskedasticity and no autocorrelation are violated

$$var\left(\varepsilon_i\right) \neq \sigma^2 \quad for \ \forall \ i \tag{2132}$$

$$covar\left(\varepsilon_i \varepsilon_j\right) \neq 0 \tag{2133}$$

The variance covariance of error is given by

$$\Omega = E\left(ee'\right) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & .. & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & .. & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & .. & \sigma_n^2 \end{bmatrix} \tag{2134}$$

$$Q'\Omega Q = \Lambda \tag{2135}$$

$$\Omega = Q\Lambda Q' = Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q' \tag{2136}$$

$$P = Q\Lambda^{\frac{1}{2}} \tag{2137}$$

$$P'\Omega P = I \quad ; \quad P'P = \Omega^{-1} \tag{2138}$$

Transform the model

$$PY = \beta PX + Pe \tag{2139}$$

$$Y^* = \beta X^* + e^* \tag{2140}$$

$Y^* = PY \quad X^* = PX \quad and \quad e^* = Pe$ $\qquad \beta_{GLS} = \left(X'P'PX\right)^{-1}\left(X'P'PY\right)$

$$\beta_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}\left(X'\Omega^{-1}Y\right) \tag{2141}$$

18.6 Estimation and Inference

Regress demand for a product (Y_i) on its own prices (X_i) as following

$$Y_i = \beta_1 + \beta_2 X_i + e_i \quad i = 1 \ ...N$$

where e_i is a randomly distributed error term for observation i.
List the OLS assumptions on error terms e_i . [5]
OLS assumptions

$$E\left(\varepsilon_i\right) = 0 \tag{2142}$$

$$E(\varepsilon_i x_i) = 0 \tag{2143}$$

$$var(\varepsilon_i) = \sigma^2 \quad for \ \forall \ i \tag{2144}$$

$$covar(\varepsilon_i \varepsilon_j) = 0 \tag{2145}$$

$$covar(\varepsilon_i X_i) = 0 \tag{2146}$$

Derive the normal equations and the OLS estimators of $\widehat{\beta}_1$ and $\widehat{\beta}_2$. [10]

$$\underset{\widehat{\beta}_1 \widehat{\beta}_2}{Min}\ S = \sum \varepsilon_i^2 = \sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_{1,i}\right)^2 \tag{2147}$$

First order conditions

$$\frac{\partial S}{\partial \widehat{\beta}_1} = 0; \frac{\partial S}{\partial \widehat{\beta}_2} = 0; \tag{2148}$$

$$\sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\right)(-1) = 0 \tag{2149}$$

$$\sum \left(Y_i - \widehat{\beta}_1 - \widehat{\beta}_2 X_i\right)(-X_i) = 0 \tag{2150}$$

$$\sum Y_i = \widehat{\beta}_1 N + \widehat{\beta}_2 \sum X_i \tag{2151}$$

$$\sum Y_i X_i = \widehat{\beta}_1 \sum X_i + \widehat{\beta}_2 \sum X_i^2 \tag{2152}$$

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix} \tag{2153}$$

A shopkeeper observed the data on quantities and prices as given in Table 1 below. What are the OLS estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ implied by these data? Is this a normal good? [15]

Table 25: Data on Quantities ane Prices

Quantities (Y_i)	5	10	15	20	25	30
Prices (X_i)	10	8	6	4	2	1

Hints: $\left[\sum X_i = 31 \quad \sum X_i^2 = 221 \quad \sum Y_i^2 = 2275; \sum Y_i = 105 \quad \sum Y_i X_i = 380\right]$;
$(X'X)^{-1} = \begin{bmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{bmatrix}$

OLS Estimators

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} N & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_i \\ \sum Y_i X_i \end{bmatrix} \tag{2154}$$

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 6 & 31 \\ 31 & 221 \end{bmatrix}^{-1} \begin{bmatrix} 105 \\ 380 \end{bmatrix}; |X'X| = \begin{vmatrix} 6 & 31 \\ 31 & 221 \end{vmatrix} = (6 \times 221) - (31 \times 31) = 365$$

$$\widehat{\beta}_1 = \frac{1}{365} \begin{vmatrix} 105 & 31 \\ 380 & 221 \end{vmatrix} = \frac{23205 - 11780}{365} = \frac{11425}{365} = 31.301 \tag{2155}$$

$$\widehat{\beta}_2 = \frac{1}{365}\begin{vmatrix} 6 & 105 \\ 31 & 380 \end{vmatrix} = \frac{2280 - 3255}{365} = \frac{-975}{365} = -2.671 \tag{2156}$$

$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i = 31.301 - 2.671X_i$
Demand decreases as price rises. $\frac{\partial Y}{\partial X} = -2.671$. Certainly this seems a normal good.
$\sum x_i^2 = \sum \left(X_i - \overline{X}\right)^2 = \sum X_i^2 - N\overline{X}^2 = 221 - 6(5.2)^2 = 58.76$
$\sum y_i^2 = \sum \left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - N\overline{Y}^2 = 2275 - 6(17.5)^2 = 437.5$
What are the total variances of e_i and Y_i? [10]

$$\widehat{e}_i = Y_i - 31.301 - (-2.671)\, X_i \tag{2157}$$

$$\widehat{e}_1 = 5 - 31.301 - (-2.671)\, 10 = 0.409 \tag{2158}$$

$$\widehat{e}_2 = 10 - 31.301 - (-2.671)\, 8 = 0.067 \tag{2159}$$

$$\widehat{e}_3 = 15 - 31.301 - (-2.671)\, 6 = -0.275 \tag{2160}$$

$$\widehat{e}_4 = 20 - 31.301 - (-2.671)\, 4 = -0.617 \tag{2161}$$

$$\widehat{e}_5 = 25 - 31.301 - (-2.671)\, 2 = -0.959 \tag{2162}$$

$$\widehat{e}_6 = 30 - 31.301 - (-2.671)\, 1 = 1.37 \tag{2163}$$

$\sum \widehat{e}_i^2 = (0.409)^2 + (0.067)^2 + (-0.275)^2 + (-0.617)^2 + (-0.959)^2 + (1.37)^2 = 3.421$
$var(\varepsilon_i) = \widehat{\sigma}^2 = \frac{\sum \widehat{\varepsilon}_i^2}{N-K} = \frac{3.421}{6-2} = 0.855$
total variance of Y_i is given by $\sum y_i^2 = \sum \left(Y_i - \overline{Y}\right)^2 = \sum Y_i^2 - N\overline{Y}^2 = 2275 - 6(17.5)^2 = 437.5$

You can do shortcut deviation method (beware that it may not guarantee that TSS is more than ESS): For instance for the same look at the following

$$\widehat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2}; \widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} \tag{2164}$$

$\sum y_i x_i = \sum Y_i X_i - \overline{Y} N \overline{X} = 380 - (5.2)\,(6)\,(17.5) = -166$
$\widehat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{-166}{58.76} = -2.825$
$\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X} = 17.5 - (-2.825)\,(5.2) = 32.19$
What are R^2 and $\overline{R}^2$? [10]
$R^2 = \frac{\sum \widehat{y}_i^2}{\sum y_i^2}$; $\overline{R}^2 = 1 - \left(1 - R^2\right) \frac{N-1}{N-K}$
$\sum \widehat{y}_i^2 = \sum y_i^2 - \sum \widehat{e}_i^2 = 437.5 - 3.421 = 434.079$
$R^2 = \frac{\sum \widehat{y}_i^2}{\sum y_i^2} = \frac{434.079}{437.5} = 0.9922$
$\overline{R}^2 = 1 - \left(1 - R^2\right) \frac{N-1}{N-K} = 1 - (1 - 0.9922)\, \frac{5}{4} = 0.9902$

Again the shortcut does not give quite the same answer
$\widehat{y}_i = \widehat{\beta}_2 x_i$
$\sum \widehat{y}_i^2 = \sum (\widehat{\beta}_2 x_i)^2 = \widehat{\beta}_2^2 \sum {x_i}^2 = (2.671)^2\,(58.76) = 419.21$
$\sum \widehat{e}_i^2 = \sum y_i^2 - \sum \widehat{y}_i^2 = 437.5 - 419.21 = 18.29$

Determine the overall significance of this model by F-test at 5 percent level of significance. [Critical value of F for df(1,4) =7.71]

$$F = \frac{RSS/(K-1)}{ESS/(N-k)} = \frac{\frac{434.079}{1}}{\frac{3.421}{4}} = \frac{434.079}{0.85525} = 507.55 \tag{2165}$$

What are the variances and standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$? [10]

Evaluate $(X'X)^{-1}$

$$(X'X)^{-1} = \begin{bmatrix} 6 & 31 \\ 31 & 221 \end{bmatrix}^{-1} = \frac{1}{\begin{vmatrix} 6 & 31 \\ 31 & 221 \end{vmatrix}} \begin{pmatrix} 221 & -31 \\ -31 & 6 \end{pmatrix}' = \frac{1}{365} \begin{pmatrix} 221 & -31 \\ -31 & 6 \end{pmatrix} = \begin{pmatrix} \frac{221}{365} & \frac{-31}{365} \\ \frac{-31}{365} & \frac{6}{365} \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} \frac{221}{365} & \frac{-31}{365} \\ \frac{-31}{365} & \frac{6}{365} \end{pmatrix} = \begin{pmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{pmatrix}$$

$$cov\left(\widehat{\beta}\right) = (X'X)^{-1}\,\widehat{\sigma}^2 = \begin{pmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{pmatrix}(0.855)$$

$$(X'X)^{-1} = \begin{pmatrix} \frac{221}{365} & \frac{-31}{365} \\ \frac{-31}{365} & \frac{6}{365} \end{pmatrix} = \begin{pmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{pmatrix}$$

$$cov\left(\widehat{\beta}\right) = (X'X)^{-1}\,\widehat{\sigma}^2 = \begin{pmatrix} 0.605 & -0.085 \\ -0.085 & 0.0164 \end{pmatrix}(0.855)$$

$$var\left(\widehat{\beta}_1\right) = (0.605)(0.855) = 0.5153$$

$$var\left(\widehat{\beta}_2\right) = (0.0164)(0.855) = 0.0140$$

$$SE\left(\widehat{\beta}_1\right) = \sqrt{var\left(\widehat{\beta}_1\right)} = \sqrt{0.5153} = 0.7178$$

$$SE\left(\widehat{\beta}_2\right) = \sqrt{var\left(\widehat{\beta}_2\right)} = \sqrt{0.0140} = 0.1183$$

Compute t-statistics and determine whether parameters $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are statistically significant at 5 percent level of significance.

[Critical value of t for five percent significance for 4 degrees of freedom is 2.776(i.e $t_{crit,0.05,4} = 2.777$)] [10]

Here the null hypothesis is that $H_0 : \beta_1 = 0$; and $\beta_2 = 0$ against alternative hypothesis $H_A : \beta_1 \neq 0$; and $\beta_2 \neq 0$

$$t\left(\widehat{\beta}_1\right) = \frac{\widehat{\beta}_1 - 0}{SE\left(\widehat{\beta}_1\right)} = \frac{31.301}{0.7178} = 43.607$$

$$t\left(\widehat{\beta}_2\right) = \frac{\widehat{\beta}_2 - 0}{SE\left(\widehat{\beta}_2\right)} = \frac{-2.671}{0.1183} = -22.5281$$

Both of these values are greater than the critical value of t. Statistical evidence suggest to H_0 and we conclude that $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are statistically significant at 5 percent level of significance.

What is the prediction of Y when X is 0.5? [5]

$\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i = 31.301 - 2.671\,(0.5) = 29.97$

What is the elasticity of demand evaluated at the mean values of Y_i and X_i? [5]

$\frac{\partial Y_i}{\partial X_i} = \beta_2 \qquad e = \frac{\partial Y_i}{\partial X_i}\frac{\overline{X}}{\overline{Y}} = \beta_2 \frac{\overline{X}}{\overline{Y}} = -2.671 \times \frac{5.2}{17.5} = -0.794$

Reformulate the model to include price of a substitute product in the model. What will happen to this estimation if these two prices are exactly correlated? [5]

let us redefine $X_{1,i}$ be the product price and $X_{2,i}$ be the price of the substitute product. Then the regression can be written as:

$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + e_i$

If there is exact collinearity like $X_{1,i} = \lambda X_{2,i}$ the OLS procedure will breakdown. In matrix approach

$$\begin{bmatrix} \sum Y_i \\ \sum X_{1,i} Y_i \\ \sum X_{2,i} Y_i \end{bmatrix} = \begin{bmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i} X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i} X_{2,i} & \sum X_{2,i}^2 \end{bmatrix} \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} \tag{2166}$$

$$|X'X| = \begin{vmatrix} N & \sum X_{1,i} & \sum X_{2,i} \\ \sum X_{1,i} & \sum X_{1,i}^2 & \sum X_{1,i} X_{2,i} \\ \sum X_{2,i} & \sum X_{1,i} X_{2,i} & \sum X_{2,i}^2 \end{vmatrix} = \begin{vmatrix} N & \lambda \sum X_{2,i} & \sum X_{2,i} \\ \lambda \sum X_{2,i} & \lambda^2 \sum X_{2,i}^2 & \lambda \sum X_{2,i}^2 \\ \sum X_{2,i} & \lambda \sum X_{2,i}^2 & \sum X_{2,i}^2 \end{vmatrix} = 0$$

It is not possible to apply the OLS procedure.

How would you decide whether demand for this product varies by gender? [5]

A new D_i dummy variable can be introduced defining it as

$D_i = \int \begin{matrix} 1 \text{ if Male} \\ 0 \text{ Female} \end{matrix}$

$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \delta D_i + e_i$

Here δ measures the difference in consumption by male against female. This is a male effect in consumption.

type I and type II errors

Elaborate on the following with relevant diagrams

	True	False
Accept	Correct	Type II error
Reject	Type I error	Correct

18.7 Distributions: Normal, t, F and chi_square

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right) \tag{2167}$$

Lognormal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(\ln x-\mu)^2}{\sigma^2} \right) \tag{2168}$$

Standard normal:

$$e \sim N(0,1) \tag{2169}$$

Any distribution can be converted to the standard normal distribution by normalization.

Chi-square: Sum of the Square of a normal distribution

$$Z = \sum_{i=1}^{k} Z_i^2 \tag{2170}$$

with k degrees of freedom.

t Distribution: ratio of normal to chi-square

$$t = \frac{Z_1}{\sqrt{Z_1/k}} \tag{2171}$$

F - distribution: ratios of two chi-square distribution with df k_1 and k_2

$$F = \frac{\sqrt{Z_1/k_1}}{\sqrt{Z_1/k_2}} \tag{2172}$$

18.8 Large Sample Theory

Probability limit

$$p\lim(\beta) = \beta \tag{2173}$$

- - Central limit theorem
  - t Distribution: ratio of normal to chi-square

$$\frac{\overline{Y} - \beta}{\sigma/\sqrt{T}} = N(0,1) \tag{2174}$$

  - Convergence in limit

$$\lim_{t \longrightarrow \infty} p\left|\widehat{\theta} - \theta\right| \leqslant \varepsilon = 1 \implies p\lim\left(\widehat{\theta}\right) = \theta \tag{2175}$$

t-distribution more accurate for finite samples but the normal distribution asymptotically approximates any other distribution according to the central limit theorem.

Probability limit of sum of two numbers is sum of probability limits

Probability limit of product of two numbers is product of probability limits

Probability limit of a function is the function of the probability limit (Slutskey theorem)

18.9 Unconstrained Optimisation

1. Consider y as a function of x_1,x_2 and x_3 as given in the following equation:

$$y = -5x_1^2 + 10x_1 + x_1x_3 - 2x_2^2 + 4x_2 + 2x_2x_3 - 4x_3^2 \tag{2176}$$

   (a) Find the optimal values of x_1,x_2 and x_3 using the first order conditions for unconstrained maximisation. Use matrix approach in your solution.

   (b) Determine whether the above solutions correspond to the minimum or the maximum point using positive or negative definite concepts of the Hessian determinants.
   Solution.
   First find the Jacobian matrix for the above equation.

$$\frac{\partial y}{\partial x_1} = -10x_1 + 10 + x_3 = 0 \tag{2177}$$

$$\frac{\partial y}{\partial x_2} = -4x_2 + 4 + 2x_3 = 0 \tag{2178}$$

$$\frac{\partial y}{\partial x_3} = x_1 + 2x_2 - 8x_3 = 0 \tag{2179}$$

Put this in the matrix format

$$\begin{pmatrix} -10x_1 & 0x_2 & x_3 \\ 0x_1 & -4x_2 & 2x_3 \\ x_1 & 2x_2 & -8x_3 \end{pmatrix} = \begin{pmatrix} -10 \\ -4 \\ 0 \end{pmatrix} \tag{2180}$$

By rearrangement:

$$\begin{pmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -10 \\ -4 \\ 0 \end{pmatrix} \tag{2181}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{pmatrix}^{-1} \begin{pmatrix} -10 \\ -4 \\ 0 \end{pmatrix} \tag{2182}$$

Define matrix notations:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}; A = \begin{pmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{pmatrix}; b = \begin{pmatrix} -10 \\ -4 \\ 0 \end{pmatrix}$$

In matrix representation off above system

$$X = A^{-1}b \tag{2183}$$

$$A^{-1} = \frac{1}{|A|} Adj\,(A) = \frac{1}{|A|} C' \tag{2184}$$

Determinant of A matrix

$$|A| = \begin{vmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{vmatrix} = -276 \tag{2185}$$

$|A| = -320 + 0 + 0 + 4 + 40 + 0 = -276$
Minors

$$M_A = \begin{pmatrix} \begin{vmatrix} -4 & 2 \\ 2 & -8 \end{vmatrix} & \begin{vmatrix} 0 & 2 \\ 1 & -8 \end{vmatrix} & \begin{vmatrix} 0 & -4 \\ 1 & 2 \end{vmatrix} \\ \begin{vmatrix} 0 & 1 \\ 2 & -8 \end{vmatrix} & \begin{vmatrix} -10 & 1 \\ 1 & -8 \end{vmatrix} & \begin{vmatrix} -10 & 0 \\ 1 & 2 \end{vmatrix} \\ \begin{vmatrix} 0 & 1 \\ -4 & 2 \end{vmatrix} & \begin{vmatrix} -10 & 1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} -10 & 0 \\ 0 & -4 \end{vmatrix} \end{pmatrix} \tag{2186}$$

Cofactor

$$C = \begin{pmatrix} \begin{vmatrix} -4 & 2 \\ 2 & -6 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 1 & -6 \end{vmatrix} & \begin{vmatrix} 0 & -4 \\ 1 & 2 \end{vmatrix} \\ -\begin{vmatrix} 0 & 1 \\ 2 & -6 \end{vmatrix} & \begin{vmatrix} -10 & 1 \\ 1 & -6 \end{vmatrix} & -\begin{vmatrix} -10 & 0 \\ 1 & 2 \end{vmatrix} \\ \begin{vmatrix} 0 & 1 \\ -4 & 2 \end{vmatrix} & -\begin{vmatrix} -10 & 1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} -10 & 0 \\ 0 & -4 \end{vmatrix} \end{pmatrix} \tag{2187}$$

$$C = \begin{pmatrix} \begin{vmatrix} -4 & 2 \\ 2 & -8 \end{vmatrix} & -\begin{vmatrix} 0 & 2 \\ 1 & -8 \end{vmatrix} & \begin{vmatrix} 0 & -4 \\ 1 & 2 \end{vmatrix} \\ -\begin{vmatrix} 0 & 1 \\ 2 & -8 \end{vmatrix} & \begin{vmatrix} -10 & 1 \\ 1 & -8 \end{vmatrix} & -\begin{vmatrix} -10 & 0 \\ 1 & 2 \end{vmatrix} \\ \begin{vmatrix} 0 & 1 \\ -4 & 2 \end{vmatrix} & -\begin{vmatrix} -10 & 1 \\ 0 & 2 \end{vmatrix} & \begin{vmatrix} -10 & 0 \\ 0 & -4 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} 28 & 2 & 4 \\ -2 & 79 & 20 \\ 4 & 20 & 40 \end{pmatrix} \tag{2188}$$

$$C' = \begin{pmatrix} 28 & -2 & 4 \\ 2 & 79 & 20 \\ 4 & 20 & 40 \end{pmatrix} \tag{2189}$$

$$A^{-1} = \frac{1}{|A|}C' = \frac{1}{-276}\begin{pmatrix} 28 & -2 & 4 \\ 2 & 79 & 20 \\ 4 & 20 & 40 \end{pmatrix} = \begin{pmatrix} -0.1014 & 0.0072 & -0.0144 \\ -0.0072 & 0.2862 & -0.07246 \\ -0.0144 & -0.07246 & 0.14493 \end{pmatrix} \tag{2190}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -0.1014 & 0.0072 & -0.0144 \\ -0.0072 & 0.2862 & -0.07246 \\ -0.0144 & -0.07246 & 0.14493 \end{pmatrix}\begin{pmatrix} -10 \\ -4 \\ 0 \end{pmatrix} = \begin{pmatrix} 1.0435 \\ 1.2174 \\ 0.4347 \end{pmatrix} \tag{2191}$$

Alternatively this system could be solved using the Cramer's rule

$$x_1 = \frac{1}{-276}\begin{vmatrix} -10 & 0 & 1 \\ -4 & -4 & 2 \\ 0 & 2 & -8 \end{vmatrix} = \frac{-288}{-276} = 1.0435 \tag{2192}$$

$$x_2 = \frac{1}{-276}\begin{vmatrix} -10 & -10 & 1 \\ 0 & -4 & 2 \\ 1 & 0 & -8 \end{vmatrix} = \frac{-336}{-276} = 1.2174 \tag{2193}$$

$$x_2 = \frac{1}{-276}\begin{vmatrix} -10 & 0 & -10 \\ 0 & -4 & -4 \\ 1 & 2 & 0 \end{vmatrix} = \frac{-120}{-276} = 0.4347 \tag{2194}$$

Properties of the Hessian Determinants

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{pmatrix} = \begin{pmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{pmatrix} \tag{2195}$$

$$H_1 = y_{11} = -10 < 0 \tag{2196}$$

$$H_2 = \begin{vmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{vmatrix} = \begin{vmatrix} -10 & 0 \\ 0 & -4 \end{vmatrix} = 40 > 0 \tag{2197}$$

$$H_3 = \begin{vmatrix} -10 & 0 & 1 \\ 0 & -4 & 2 \\ 1 & 2 & -8 \end{vmatrix} = -320 + 0 + 0 + 4 + 40 + 0 = -276 = 40 > 0 \tag{2198}$$

18.9.1 Nobel Laureates in Economics

2021: David Card, Joshua D. Angrist and Guido W. Imbens
2020: Paul Milgrom and Robert Wilson
2019: Abhijit Benerji, Esther Duflo and Michael Kremer
2018: Paul Romer and William Nordhaus
2017: Richard H. Thale
2016: Oliver Hart and Bengt Holmstrom
2015: Angus Deaton
2014: Jeane Tirole
2013: Hansen L P, E. Fama and R. Shiller
2012: Alvin E. Roth and Lloyd S. Shapley
2011: Thomas J. Sargent, Christopher A. Sims
2010: Peter A. Diamond, Dale T. Mortensen, Christopher A. Pissarides
2009: Elinor Ostrom, Oliver E. Williamson
2008: Paul Krugman
2007: Leonid Hurwicz, Eric S. Maskin, Roger B. Myerson
2006: Edmund S. Phelps
2005: Robert J. Aumann, Thomas C. Schelling
2004: Finn E. Kydland, Edward C. Prescott
2003: Robert F. Engle III, Clive W.J. Granger
2002: Daniel Kahneman, Vernon L. Smith
2001: George A. Akerlof, A. Michael Spence, Joseph E. Stiglitz
2000: James J. Heckman, Daniel L. McFadden
1999: Robert A. Mundell
1998: Amartya Sen
1997: Robert C. Merton, Myron S. Scholes
1996: James A. Mirrlees, William Vickrey
1995: Robert E. Lucas Jr.
1994: John C. Harsanyi, John F. Nash Jr., Reinhard Selten
1993: Robert W. Fogel, Douglass C. North
1992: Gary S. Becker
1991: Ronald H. Coase
1990: Harry M. Markowitz, Merton H. Miller, William F. Sharpe
1989: Trygve Haavelmo
1988: Maurice Allais

1987: Robert M. Solow
1986: James M. Buchanan Jr.
1985: Franco Modigliani
1984: Richard Stone
1983: Gerard Debreu
1982: George J. Stigler
1981: James Tobin
1980: Lawrence R. Klein
1979: Theodore W. Schultz, Sir Arthur Lewis
1978: Herbert A. Simon
1977: Bertil Ohlin, James E. Meade
1976: Milton Friedman
1975: Leonid Vitaliyevich Kantorovich, Tjalling C. Koopmans
1974: Gunnar Myrdal, Friedrich August von Hayek
1973: Wassily Leontief
1972: John R. Hicks, Kenneth J. Arrow
1971: Simon Kuznets
1970: Paul A. Samuelson
1969: Ragnar Frisch, Jan Tinbergen
http://www.nobelprize.org/nobel_prizes/economics/laureates/