

# Quantile Stochastic Frontier Models with Endogeneity

Mike G. Tsionas, \*\* A. George Assaf , \* Athanasios Andrikopoulos<sup>†</sup>

## Abstract

In this paper, we extend Jradi et al. (2019). First, we use the asymmetric Laplace distribution, which is a more reasonable assumption in quantile models. Second, we address the issue of statistical inference for the optimal quantile. Finally, we allow for endogeneity in quantile stochastic frontier models. The new formulation is implemented in a Bayesian framework using Markov Chain Monte Carlo. We employ the celebrated Philippine rice data as in Jradi et al. (2019). Jradi et al. (2019) did not provide efficiency measures, which, in our framework, is straightforward to do.

**Keywords:** Quantile estimation; Stochastic frontier models; Efficiency, Bayesian analysis; Markov Chain Monte Carlo.

**JEL codes:** C11, C13.

---

Corresponding Author: \*Albert. George Assaf. Isenberg School of Management, University of Massachusetts-Amherst, 90 Campus Center Way, 204A Flint Lab, Amherst, MA, 01003, United States, [assaf@isenberg.umass.edu](mailto:assaf@isenberg.umass.edu).\*\* Mike G. Tsionas. Lancaster University Management School, LA1 4YX, UK., [m.tsionas@lancaster.ac.uk](mailto:m.tsionas@lancaster.ac.uk).<sup>†</sup>Athanasios Andrikopoulos. Hull University Business School, University of Hull, Hull HU6 7RX, UK., [A.Andrikopoulos@hull.ac.uk](mailto:A.Andrikopoulos@hull.ac.uk).

**Acknowledgments:** We are grateful to an anonymous reviewer for several useful remarks on an earlier version.

# 1. Introduction

Quantile estimation of stochastic frontier models has received a lot of attention in the literature (Behr, 2010; Bernini et al., 2004; Knox et al., 2007; Kumbhakar et al., 2019, Liu et al., 2008) including the fundamental papers of Jradi and Ruggiero (2019). As Jradi, Parmeter and Ruggiero (2019) write: “An obvious appeal of quantile regression is that it is known to provide a more complete picture of a conditional distribution [...] and is a robust alternative to regression based methods. Whereas the conditional quantile estimator is determined through minimization of the “check” function (Koenker and Bassett, 1978) defined for a particular quantile, the median say, the ordinary least squares estimator stems from minimization of the sum of squared errors making it susceptible to outliers.” (Jradi, Ruggiero and Parmeter, 2019, p. 15).

In this paper, we provide an alternative to Jradi, Parmeter and Ruggiero (2019). Jradi, Parmeter and Ruggiero (2019) provide an estimate of the optimal quantile  $\hat{\tau}^*$  provided we have a normal-half normal distribution. In our case, conditional on inefficiency, the two-sided error component follows an asymmetric Laplace distribution (ALD), which is more consistent with the quantile regression formulation, and then inefficiency is assigned a given distribution. As Jradi, Ruggiero and Parmeter (2019, page 17) write: “One may ask why we elect to determine  $\hat{\tau}^*$  by minimizing the distance over a fixed grid of selected  $\tau$ s. [...] This is an important issue and one that we leave for future research”. As they also write in their footnote 6, “An important issue for future work is to examine the statistical properties of  $\hat{\tau}^*$ ”.

Our novelty relative to Jradi, Parmeter and Ruggiero (2019) is that the one-sided error, conditionally on inefficiency, follows an ALD, so we do not have a normal-half normal specification. More specifically, we have what might be called an ALD–half normal specification (see also Tsionas, 2020). The issue of distributional properties of  $\hat{\tau}^*$  is addressed directly and we can, in fact, provide the marginal posterior of the optimal quantile. Our second novelty is that although Jradi, Ruggiero and Parmeter (2019) did not provide efficiency measures, in our framework this is straightforward to do. Third, we allow for endogeneity which we know it is quite important when estimating production functions (in the quantile framework or otherwise). Finally, we feel it is important to mention that a big caveat that should be mentioned about JPR is that their paper does not engage in quantile estimation of the stochastic frontier model per se, but attempts to find the quantile that is consistent with the exact location of the

frontier. Given that the frontier is a well conceived concept, this is quite different than looking at a corresponding quantile in the conditional distribution of output. Both are equally important, but different and JPR really only look at the first. Our approach looks at the second. For related work, see Horrace and Parmeter (2018).<sup>1</sup>

## 2. Model

In what follows we first present the stochastic frontier model with ALD errors without taking account of endogeneity. In Section 2.2 we take up the issue of endogeneity which is common in applied production studies.

### 2.1. Not accounting for endogeneity

Suppose

$$y_i = x'_i \beta + v_i - u_i, \quad i = 1, \dots, n, \quad (1)$$

where  $x_i$  is a  $k \times 1$  vector of regressors<sup>2</sup>,  $\beta$  is a  $k \times 1$  vector of parameters,  $v_i$  is a two-sided error terms, and  $u_i$  is a non-negative error component representing technical inefficiency. Let us take inefficiency given for the moment. The standard quantile regression problem is based on the asymmetric Laplace distribution (ALD) for  $v_i$  and provides the maximum likelihood estimator when

$$y_i | x_i, u_i \sim ALD(x'_i \beta - u_i, \sigma_v, \tau), \quad i = 1, \dots, n, \quad (2)$$

where  $\sigma_v > 0$  is a scale parameter,  $\tau \in (0,1)$  denotes quantile, and  $ALD(\mu, \sigma_v, \tau)$  has density which is given by  $f(e) = \frac{\tau(1-\tau)}{\sigma_v} \exp \left\{ -\frac{e(\tau - \mathbb{1}(e < 0))}{\sigma_v} \right\}$ . Clearly, we can express this model as:

$$y_i = x'_i \beta_\tau + v_i - u_i, \quad i = 1, \dots, n, \quad (3)$$

where  $v_i | x_i, u_i \sim ALD(0, \sigma_v, \tau)$ ,  $i = 1, \dots, n$ . We also assume  $u_i \sim N_+(0, \sigma_u^2)$ ,  $i = 1, \dots, n$ . Our problem in this paper is to investigate the endogeneity of  $x_i$ , viz. when  $x_i$  orthogonal to or independent of  $v_i$ .

An alternative and, perhaps, more useful representation is:

$$y_i = x'_i \beta + \sigma_v A(\tau) \xi_i + v_i - u_i, \quad i = 1, \dots, n, \quad (4)$$

---

<sup>1</sup> We would like to thank an anonymous reviewer for raising this point.

<sup>2</sup>For simplicity we focus on the cross-sectional case but the same analysis applies to panel data as we will see in our application.

where

$$v_i | \xi_i \sim N(0, \sigma_v^2 B(\tau) \xi_i), \quad i = 1, \dots, n, \quad (5)$$

where  $A(\tau) = \frac{1-2\tau}{\tau(1-\tau)}$ ,  $B(\tau) = \frac{2}{\tau(1-\tau)}$  and  $\xi_i$  follows a standard exponential distribution, see for example Kuzobowski and Podgorski (2000) and Lum and Gelfand (2012). This representation is different compared to Tsionas (2020) which leads to differences in MCMC formulations. In the absence of endogeneity, and in a Bayesian framework it is easy to think of  $\tau$  as part of the parameter vector in which case a prior  $p(\tau | \beta, \sigma_v, \sigma_u)$  is assigned, say  $p(\tau | \beta, \sigma_v, \sigma_u) \propto \text{const.}$ , and  $\tau \in (0,1)$ . This prior is non-informative. Assuming  $p(\beta, \sigma_v, \sigma_u) \propto \sigma_v^{-1} \sigma_u^{-1}$ , viz. a non-informative prior, the posterior becomes:

$$p(\theta, u, \tau | D) \propto [\tau(1-\tau)]^n \sigma_v^{-(n+1)} \sigma_u^{-(n+1)} \exp \left\{ -\frac{\sum_{i=1}^n (y_i - x_i' \beta + u_i) (\tau - \mathbb{1}(y_i - x_i' \beta + u_i < 0))}{\sigma_v} - \frac{\sum_{i=1}^n u_i^2}{2\sigma_u^2} \right\} p(\theta). \quad (6)$$

## 2.2 Accounting for endogeneity

To account for endogeneity, in the presence of instruments  $z_i \in \mathbb{R}^{d_z}$  we use the following “reduced form”:

$$x_i = \Pi z_i + v_{i0}, \quad i = 1, \dots, n, \quad (7)$$

where  $\Pi$  is a  $k \times d_z$  matrix, and

$$V_i = \begin{bmatrix} v_i \\ v_{i0} \end{bmatrix} | \xi_i \sim N_{k+1}(0, A(\tau) \xi_i \Sigma), \quad i = 1, \dots, n, \quad (8)$$

subject to the restriction that  $\sigma_{11} = \sigma_v^2$ , where  $\Sigma$  is a  $(k+1) \times (k+1)$  covariance matrix. Notice that dependence between  $v_i$  and  $v_{i0}$  is modeled via both the general form of  $\Sigma$  and  $\xi_i$  as well. Jointly, (5) and (8), *unconditionally* on  $\xi_i$  form a multivariate ALD. The details of our Markov Chain Monte Carlo (MCMC) are presented in the Appendix. Relative to Tsionas (2020) the specification in (5) and (8), becomes quite different once endogeneity is taken into account as the reduced form now depends on  $\tau$  and, in effect, the conditional distribution of  $x_i$ s shifts with  $\tau$ . This fact allows straightforward generalizations of the present model to the case of input-

oriented or output-oriented distance functions.

### 3. Application

We use the same data as in Jradi, Parmeter and Ruggiero (2019), viz. the celebrated rice farming data collected in the Philippines, see Coelli et al. (2005) and Parmeter et al. (2019). The data are composed of 43 farmers observed annually for eight years. Even though the data constitutes a panel, we ignore this as in Jradi, Parmeter and Ruggiero (2019).

The output variable ( $y$ ) is tonnes of freshly threshed rice. The input variables are area ( $x_1$ ) of planted rice ( $x_2$ ), total labor ( $x_3$ ) used (man-days of family and hired-labor) and fertilizer ( $x_4$ ). Following Rho and Schmidt (2015) we normalize the inputs and output at their respective means prior to taking the log transformation. This implies that we can sum the coefficients on the linear terms of the translog production frontier to calculate returns to scale (RTS) at the means.

Endogeneity plagues applied econometrics practice and, more often than not, it is a serious issue. In the context of production functions, for example, it is known that inputs are endogenous, at least since Zellner, Kmenta and Dreze (1966). Input endogeneity holds under most behavioral assumptions for the firm and, of course, when noise/shocks are known by the firm but not the economist as in this case, input decisions necessarily depend on the realization of shocks. Moreover, as rice production depends a lot on weather and other types of shocks, it is highly unlikely that inputs are exogenous or predetermined so considering endogeneity is extremely important in this application.

Access to the posterior is provided via Markov Chain Monte Carlo (MCMC) using 150,000 iterations, omitting the first 50,000 to mitigate possible start-up effects. To check convergence we have used ten different chains starting from different initial conditions. Drawing MCMC for  $\beta, \sigma_v, \sigma_u$  is straightforward. For  $u_i$ s and  $\tau$  this is not the case but we provide the details in the Appendix.

Our parameter estimates are reported in Table 1, although they do not have a structural interpretation, to compare with Jradi, Parmeter and Ruggiero (2019). Our marginal posterior density of  $\tau$  is reported in Fig. 1.

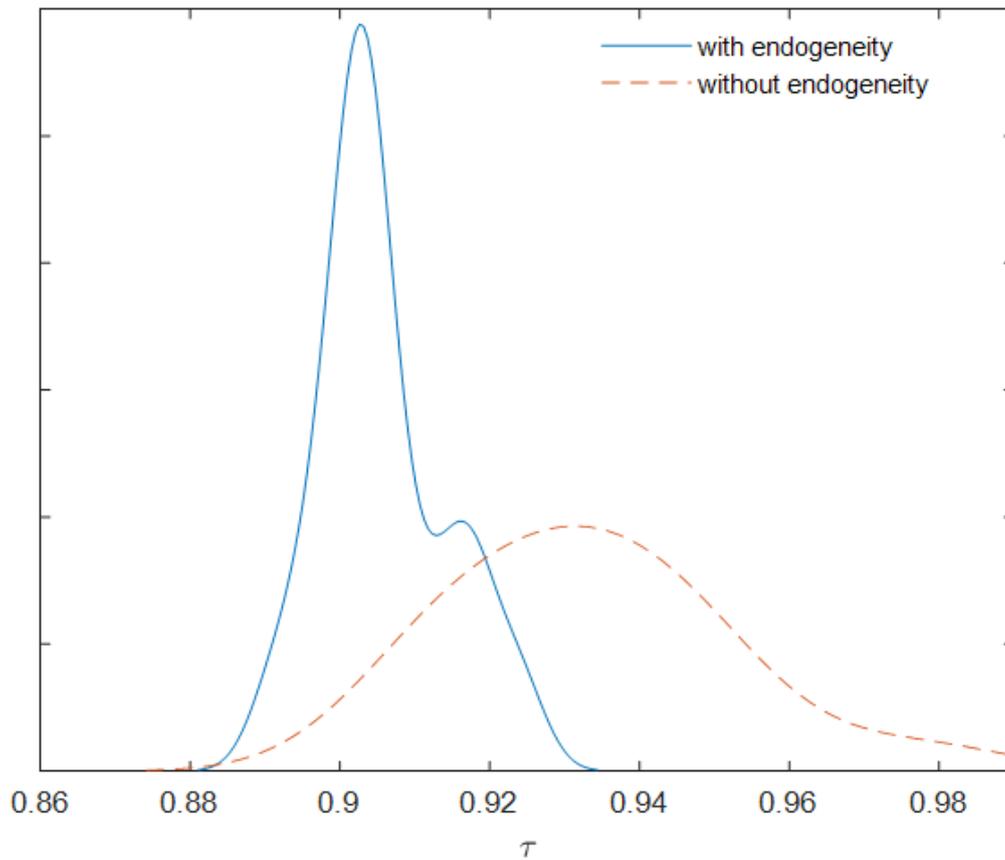


Fig. 1 Marginal posterior density for  $\tau$

Interestingly, the marginal posterior is bimodal with two modes near 0.90 and 0.92. The posterior mean is 0.906 with posterior standard deviation 0.0085, and the posterior median is 0.903.

**Table 1. Posterior moments**

---

$\beta_1$	0.372 (0.032)
$\beta_2$	0.303 (0.044)
$\beta_3$	0.322 (0.017)
$\beta_{11}$	-0.117 (0.016)
$\beta_{12}$	-0.255 (0.019)
$\beta_{13}$	0.135 (0.020)
$\beta_{22}$	0.214 (0.019)
$\beta_{23}$	-0.181 (0.025)
$\beta_{33}$	-0.122 (0.044)
trend	0.005 (0.002)
$\lambda$	3.403 (0.019)
$\sigma^2$	0.127 (0.025)
$\tau$	0.912 (0.014)

---

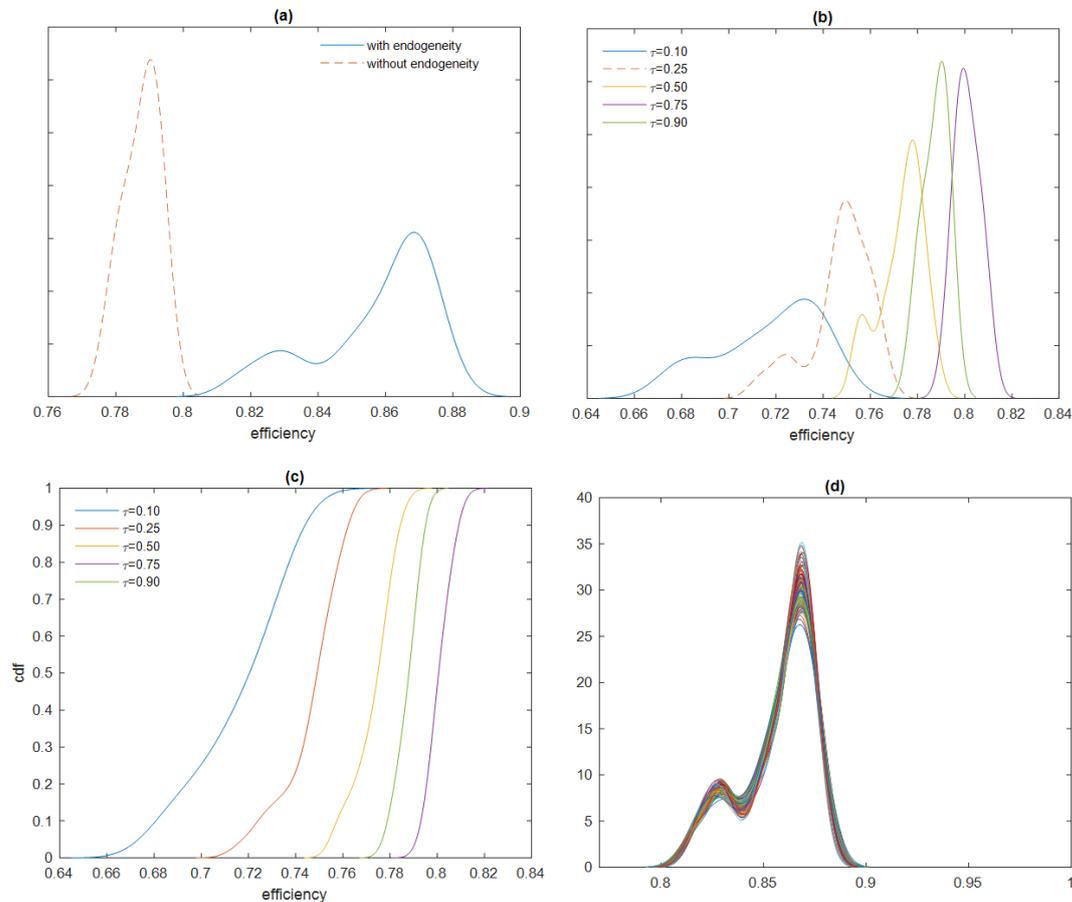
Notes: The coefficients are the same as in Eq. (7) in Jradi, Parmeter and Ruggiero (2019). “trend” corresponds to their  $\theta$ . Moreover,  $\sigma^2 = \sigma_v^2 + \sigma_u^2$ ,  $\lambda = \frac{\sigma_u}{\sigma_v}$  and  $\tau$  is the quantile parameter.

Our returns to scale (RTS) measure averages<sup>3</sup> to 0.997 (with posterior s.d. 0.032) compared to 0.963 in Jradi, Parmeter and Ruggiero (2019). To the best of our knowledge Jradi, Parmeter and Ruggiero (2019) did not provide efficiency measures.

---

<sup>3</sup>If the  $x$ s are in deviations from means, then the sum of  $\beta_j$ s provides average RTS.

In a Bayesian setting (in) efficiency can be estimated easily using the MCMC draws for  $u_i$ s. Their sample distribution is provided in Fig. 2. The mean and standard deviation of efficiency estimates are, respectively, 0.873 and 0.0415.



**Fig. 2. Sample densities of efficiency**

Estimates of efficiency are different in panel (a) of Fig. 2, the intuition being that inconsistent parameter estimates, derived when endogeneity is ignored, have a large impact on such estimates. The fact that such estimates are *quite* different is interpreted as direct evidence that the amount of endogeneity is substantial. To the extent, therefore, that parameter estimates of the frontier are quite different, implies that efficiency estimates are quite different as well. From panel (a) the differences are, indeed, substantial.

The point that  $\tau$  can be estimated and, therefore, it has a marginal posterior distribution, has been overlooked in the literature and, for the most part, applied researchers would be interested in something like Fig. 2 which shows how inefficiency varies with a *given* value of  $\tau$ . This point of view is not necessarily correct. In panel

(a) of Figure 1, we present the marginal posterior of  $\tau$  when uncertainty with respect to this parameter is taken into account (that is when  $\tau$  is integrated out of the joint posterior of inefficiency,  $\tau$  and other parameters). So, the evidence in panel (b) of Fig. 2 would be more appropriate for researchers that want to look into efficiency distributions for a given quantile, but if uncertainty with respect to the particular quantile is important, then empirical researchers need to look at panel (a) of Fig. 2.

In panel (c) of the same Figure, we present cumulative distribution functions (cdf) corresponding to different quantiles. From this evidence, it is clear that we do not have the “quantile crossing problem”. In other applications, this could happen and more research is needed to address this issue.

Finally, in panel (d) we present sample densities of inefficiency estimates similar to those reported in panel (a) of Fig. 1, when we consider alternative priors for the quantile parameter, viz.  $p(\tau|\beta, \sigma_v, \sigma_u) \propto \tau^{\underline{a}-1}(1-\tau)^{\underline{b}-1}$ , viz. a beta distribution, where the prior parameters  $\underline{a}$ ,  $\underline{b}$  are randomly selected (i.e. uniformly distributed) in the interval 0 to 10. MCMC is implemented again 10,000 times and 100 representative inefficiency densities are presented in panel (d) of Figure 2. These densities are quite similar to the one reported in panel (a) of Fig. 1 as well as to one another so, we do not have sensitivity with respect to a reasonable prior.<sup>4</sup>

## Concluding remarks

In this paper we take up the issue of quantile estimation in stochastic frontier models using the asymmetric Laplace distribution (ALD). The ALD is the distribution of the two-sided error, conditionally on technical inefficiency. In turn, the one-sided error term that represents inefficiency is assigned a half-normal distribution. We believe that, relative to the current state of the art, we improve quantile estimation of stochastic frontier models in three ways. First, the ALD is well-suited as a distribution of the two-sided error term in the sense that this is closely associated with the ALD interpretation of quantile regression in regression models. Second, our MCMC techniques provide directly (in) efficiency estimates utilizing fully our assumptions. Third, we successfully address the issue of statistical inferences for the optimal quantile and, additionally, we

---

<sup>4</sup>We are grateful to a reviewer for pointing out the necessity of this exercise.

address formally the endogeneity problem. In terms of future research, it would be quite interesting to see how the no-quantile-crossing property can be imposed (see Wang et al., 2014, Tsionas, 2020).

## References

- Behr, A., 2010. Quantile regression for robust bank efficiency score estimation. *European Journal of Operational Research* 200, 568–581.
- Bernini, C., Freo, M., Gardini, A., 2004. Quantile estimation of frontier production function. *Empirical Economics* 29, 373–381.
- Coelli, T.J., Rao, D.P., O'Donnell, C.J., Battese, G.E., 2005. *An Introduction to Efficiency and Productivity Analysis*. Springer, New York, New York.
- Horrace, W. C., Parmeter, C. F., 2018. A Laplace Stochastic Frontier Model. *Econometric Reviews*, 37 (3), 260-280.
- Jradi, S., Parmeter, C.F., Ruggiero, J., 2019. Quantile estimation of the stochastic frontier model. *Economics Letters* 182, 15–18.
- Jradi, S., Ruggiero, J., 2019. Stochastic data envelopment analysis: A quantile regression approach to estimate the production frontier. *European Journal of Operational Research* 278 (2), 385–393.
- Knox, K.J., Blankmeyer, E.C., Stutzman, J.R., 2007. Technical efficiency in Texan nursing facilities: A stochastic production frontier approach. *Journal of Economics and Finance* 31 (1), 75–86.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46 (1), 33–50.
- Kumbhakar, S.C., Parmeter, C.F., Zelenyuk, V., 2019. Stochastic frontier analysis: Foundations and advances II. In: Ray, S., Chambers, R., Kumbhakar, S.C. (Eds.), *Handbook of Production Economics*. Springer (forthcoming).
- Kuzobowski, T. J. and Podgorski, K. (2000). A Multivariate and Asymmetric Generalization of Laplace Distribution. *Computational Statistics*, 15 (4), 531–540.
- Liu, C., Laporte, A., Ferguson, B.S., 2008. The quantile regression approach to efficiency measurement: Insights from Monte Carlo simulations. *Health Economics* 17, 1073–1087.
- Lum, K., and Gelfand, A. E. (2012). Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process. *Bayesian Analysis* 7 (2), 235–258.

- Parmeter, C.F., Wan, A.T.K., Zheng, X., 2019. Model averaging estimators for the stochastic frontier model. *Journal of Productivity Analysis* 51 (2–3), 91–103.
- Tsionas, M. G., 2020. Quantile Stochastic Frontiers. *European Journal of Operational Research*, forthcoming.
- Rho, S., Schmidt, P., 2015). Are all firms inefficient?. *Journal of Productivity Analysis* 43 (3), 327–349.
- Wang, Y., Wang, S., Dang, C., Ge, W., 2014. Nonparametric quantile frontier estimation under shape restriction. *European Journal of Operational Research* 232, 671–678.
- Zellner, A., Kmenta, J., and Dreze, J., 1966, Specification and Estimation of Cobb-Douglas Production Function Models. *Econometrica* 34, 784–795.

## Appendix. Technical Appendix

In this Appendix, we show how to implement MCMC. The major difficulty in implementing numerical inference by MCMC is that in (8) the covariance matrix is  $\Sigma_i \equiv A(\tau)\xi_i\Sigma$  which depends on both  $\tau$  and  $\xi_i$ . Suppose  $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_1' \\ \sigma_1 & \Sigma_{11} \end{bmatrix}$  where  $\Sigma_{11}$  is  $k \times k$  and  $\sigma_1$  is  $k \times 1$ .

Using results from partitioned inversion we have:

$$\begin{aligned} & \Sigma_i^{-1} \\ &= [A(\tau)\xi_i]^{-1} \begin{bmatrix} \frac{1}{\sigma_{11} - \sigma_1'\Sigma_{11}^{-1}\sigma_1} & \cdot \\ -\Sigma_{11}^{-1}\sigma_1(\sigma_{11} - \sigma_1'\Sigma_{11}^{-1}\sigma_1)^{-1} & (\Sigma_{11} - \sigma_1'\Sigma_{11}^{-1}\sigma_1)^{-1} \end{bmatrix}. \end{aligned} \quad (\text{A.1})$$

In turn we can write the augmented posterior distribution as follows:

$$\begin{aligned} & p(\theta, \tau, \Sigma, \{u_i\}_{i=1}^n, \{\xi_i\}_{i=1}^n | D) \propto \\ & \sigma_v^{-(n+1)} \prod_{i=1}^n [A(\tau)\xi_i]^{-1} |\Sigma|^{-n/2} \cdot p(\theta, \tau, \sigma) \cdot \exp \left( -\frac{Q}{2} - \frac{1}{2\sigma_u^2} \sum_{i=1}^n u_i^2 \right) \cdot \prod_{i=1}^n I(0 < \xi_i < 1), \end{aligned}$$

where

$$\begin{aligned}
Q = & \left( \prod_{i=1}^n [A(\tau)\xi_i]^{-1} \right) [y_i - x'_i\beta - \sigma_v A(\tau)\xi_i + u_i \\
& \vdots x_i \\
& - \Pi z_i]' \Sigma^{-1} \begin{bmatrix} y_i - x'_i\beta - \sigma_v A(\tau)\xi_i + u_i \\ x_i - \Pi z_i \end{bmatrix},
\end{aligned} \tag{A.2}$$

$$\text{and } \Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_{11} - \sigma_1' \Sigma_{11}^{-1} \sigma_1} & \cdot \\ -\Sigma_{11}^{-1} \sigma_1 (\sigma_{11} - \sigma_1' \Sigma_{11}^{-1} \sigma_1)^{-1} & (\Sigma_{11} - \sigma_1' \Sigma_{11}^{-1} \sigma_1)^{-1} \end{bmatrix}.$$

To simplify the computations as much as possible we adopt the following strategy:

1. Draw a candidate  $\beta$  from  $N(b, V_b)$  where  $b = (X'X)^{-1}X'[y - \sigma_v A(\tau)\xi + u]$ ,  $V_b = h\sigma_v^2(\prod_{i=1}^n [A(\tau)\xi_i]^{-1})(X'X)^{-1}$  where  $h$  is a positive constant. We use a Metropolis-Hastings criterion to determine whether the candidate is accepted, otherwise we set the new MCMC draw to its previous value. Regarding  $h$ , we determine it so that during the burn-in phase so that, approximately, 25% of candidates are accepted.
2. We draw  $\sigma_v, \tau$ , and  $\{\xi_i\}_{i=1}^n$  using a random-walk Metropolis Hastings algorithm.
3. We draw  $\{u_i\}_{i=1}^n$  from  $N_+(0, h_o\sigma_u^2)$  and then we use a Metropolis-Hastings criterion to determine whether the candidate is accepted. Regarding  $h_o$ , we determine it so that during the burn-in phase so that, approximately, 25% of candidates are accepted.