Improving the predictability of the oil–US stock nexus: the role of macroeconomic variables

Afees A. Salisu[1][2], Raymond Swaray[3], Tirimisiyu F. Oloko[4]

Abstract

In this study, we revisit the oil–stock nexus by accounting for the role of macroeconomic variables and testing their in-sample and out-of-sample predictive powers. We follow the approaches of Lewellen (2004) and Westerlund and Narayan (2015), which were formulated into a linear multi-predictive form by Makin et al. (2014) and Salisu et al. (2018) and a nonlinear multi-predictive model by Salisu and Isah (2018). Thereafter, we extend the multi-predictive model to account for structural breaks and asymmetries. Our analyses are conducted on aggregate and sectoral stock price indexes for the US stock market. Our proposed predictive model, which accounts for macroeconomic variables, outperforms the oil-based single-factor variant in forecasting aggregate and sectoral US stocks for both in-sample and out-of-sample forecasts. We find that it is important to account for structural breaks in our proposed predictive model, although asymmetries do not seem to improve predictability. In addition, we show that it is important to pre-test the predictors for persistence, endogeneity, and conditional heteroscedasticity, particularly when modeling with high-frequency series. Our results are robust to different forecast measures and forecast horizons.

Keywords: Oil price; Sectoral US stock; Inflation; Output; Interest rate; Forecast evaluation

JEL Classification: G11, Q43

---

[1] Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City,Vietnam.

[2] Faculty of Business Administration, Ton Duc Thang University, Ho Chi Minh City, Vietnam. Email: afees.adebare.salisu@tdt.edu.vn

[3] Economics Subject Group, University of Hull Business, University of Hull, Cottingham Road, UK. Email: r.swaray@hull.ac.uk.

[4] Centre for Econometric & Allied Research, University of Ibadan, Nigeria. Email: tf.oloko@cear.org.ng.

Improving the predictability of the oil–US stock nexus: the role of macroeconomic variables

1 Introduction

The incentive to link stock returns to oil price is partly derived from the pioneering work of Hamilton (1983), which shows that energy price increases may account for much of macroeconomic performance. Hamilton's (1983) work formed the basis of an ample literature on the relationships between energy and financial markets (Arouri and Rault, 2010; Ciner, 2013; Kang et al, 2014; Sadorsky, 2014; Narayan and Gupta, 2015; Bondia, 2016; Reboredo et al.; 2017; Swaray and Salisu, 2018). Notwithstanding the vastness of the empirical literature and the data on the oil–stock relationship or the sophistication of methodological approaches (for a review of the literature on the oil–stock nexus, see Salisu and Oloko, 2015; Salisu and Isah, 2017; Smyth and Narayan, 2018), gaps remain. For example, Smyth and Narayan (2018) highlight some notable gaps in previous studies that directly relate oil to the stock market, eschewing macroeconomic fundamentals. The few exceptions include studies that control for one or more macro variables (Basher et al., 2012; Conrad et al. 2014; Liu et al., 2015; Chiang and Hughen, 2017; Roubaud and Arouri, 2018).

The works of Narayan and Sharma (2014), Narayan and Gupta (2015), and Phan et al. (2015) are notable for evaluating both in-sample and out-of-sample forecasts for oil–stock predictability, albeit with the exclusion of macroeconomic variables. Thus, there is a dearth of studies that allow for the role of macroeconomic variables in predictive models for oil–stock relationships, particularly for out-of-sample predictability. The traditional determinants of stock returns—including, but not restricted to, dividend/price ratio, dividend payout ratio, earning price ratio, short-term interest rate variable (such as the three-month Treasury Bill Rate) and long-term yield—are often used in the modeling framework (see Liu et al., 2015; Wu and Lee, 2015). The consideration of macroeconomic variables in oil–stock relationships in this paper is motivated from perspectives of theory, investment, economics and policy, and methodology.

From a theoretical standpoint, arbitrage pricing theory (APT) assumes that financial assets, particularly financial securities such as stocks, can be influenced by macroeconomic factors such as inflation, real level of economic activity, interest rate, and exchange rate. These macro variables are seen as sources of systematic risk that affect the entire stock market, unlike the unsystematic risks that are stock-specific. The work of Huang et al. (1996) and Jones and Kaul (1996), with

further contributions from Faff and Brailsford (1999) and Sadorsky (1999), established that oil prices can affect stock prices directly, by impacting on future cash flows, or indirectly, through impact on the interest rate used to discount future cash flows (see Basher et al., 2012). Since the pass-through is through the interest rate, which is also susceptible to inflationary pressure, it becomes necessary to control for interest rates and related macro variables in the analysis of the oil–stock link. In addition, for a large, open economy such as the US, the stock market cannot be immune from the influence of macroeconomic dynamics; limiting the predictability of its stock returns to oil price may therefore produce undesirable results.

From the investment perspective, information about the influence of macroeconomic variables on stock returns is useful for portfolio diversification purposes and therefore allows investors to be better informed when making investment decisions. For instance, during a period of high inflation, investors may invest in inflation-resistant economic sectors. Given high interest rates, investors may disinvest from stocks into newly issued bonds. However, exchange rate depreciation enhances trade competitiveness, boosting production and profits and, by extension, stock returns (Salisu and Oloko, 2015). From the economic and policy perspective, knowledge of the impact of macroeconomic variables on the oil–stock relationship helps policymakers understand the resilience of the economy to oil price shock and its spillover effects on the stock market. This helps to identify appropriate policy actions to take within the domestic economy in the event of exogenous shocks, particularly those due to oil price. In other words, adequate monitoring of the response of macroeconomic indicators to oil price movements is important for effective policy decision-making and for achieving superior macroeconomic outcomes. This is one of the justifications for meetings of the monetary policy committees of central banks, who play a primary role in tracking how macroeconomic indicators respond to internal and external shocks.

From the methodological perspective, allowing for macroeconomic dynamics when modeling the oil–stock nexus will enhance forecast performance only when the appropriate methodology is employed (see also Phan et al., 2015). Some studies hold the view that predictive regression models for stock returns that consider a wide range of macroeconomic predictors will perform more poorly than the historical average forecast, particularly when the analysis is extended to out-of-sample forecasts (see Bossaerts and Hillion, 1999; Goyal and Welch, 2003; Brennan and Xia, 2005; Butler et al., 2005; Ang and Bekaert, 2007; Welch and Goyal, 2008). However, recent studies, both those that are not oil-based (see Bannigidadmath and Narayan, 2015;

Narayan and Bannigidadmath, 2015; Devpura et al., 2018) and those that are oil-based (see Narayan and Gupta, 2015; Phan et al., 2015) upturned the results of earlier studies by using the approach of Lewellen (2004), which was later extended by Westerlund and Narayan (2012, 2015) (hereafter, the WN approach). The underlying theoretical framework for the implementation of this approach is presented in the Appendix. The WN approach accounts for the salient features of predictors, such as endogeneity, conditional heteroscedasticity, and persistence effects. However, the WN approach is single-factor based, and since our study allows for macroeconomic information in the predictive model for the oil–stock connection, the multi-factor framework is employed (Makin et al., 2014; Salisu and Isah, 2018; Salisu et al., 2018).

This study will be the first to formulate such a multi-predictive model to evaluate in-sample and out-of-sample predictability for stock returns; it is evident that multi-predictive models proposed by Makin et al. (2014), Salisu and Isah (2018), and Salisu et al. (2018) are not suitable for forecasting stock returns. Moreover, we extend the multi-predictive model to account for nonlinearities (see also Kilian and Park, 2009; Filis et al., 2011; Asteriou and Bashmakova, 2013; Jouini, 2013; Soucek and Todorova, 2013; Salisu and Oloko, 2015; Bouri et al., 2016; Kang et al., 2016; Li et al., 2017; Salisu and Isah, 2017) and structural breaks (see also Salisu and Fasanya, 2013; Salisu and Oloko, 2015; Salisu et al., 2017), which tend to influence the oil–stock nexus.

Thus, our analyses are conducted in the following distinct ways. First, we examine the in-sample predictability of oil price and macroeconomic indicators, namely real economic activity (Kilian 2009; Kilian and Park, 2009; Salisu et al., 2017), inflation (Devpura et al., 2018), and interest rate (Devpura et al., 2018). We are careful not to have too many predictors in the model, something that will likely lead to in-sample over-fitting problems (Liu et al., 2015). Thus, the choice of macro variables is motivated by the evidence in the literature with regard to the US, which is the focus of this study. Second, in order to test for the need to account for macro variables, we compare the forecast performance (both the in-sample and out-of-sample forecasts) of the single-factor oil-based predictive model for stock returns with the multi-factor variant that allows for the same. Third, we allow for oil price asymmetries in the multi-factor predictive model, and we test whether their inclusion will offer better in-sample and out-of-sample forecast results than the multi-factor model that ignores them. Fourth, we allow for structural breaks in the multi-factor predictive model, and we test whether such adjustment to the model will matter for the predictability of stock returns. Fifth, all the analyses are conducted for both aggregate and sectoral

4

stock price indexes for the US stock market. This offers a comprehensive assessment as regards the extent to which the information contained in oil and macroeconomic variables can be exploited to produce accurate forecast for US stock returns, both at the aggregate and sectoral levels. Finally, in order to test the significance of our economic models (i.e., the predictive models that account for the role of macroeconomic variables), we compare their forecast performance with time series models such as the fractionally integrated autoregressive moving average (ARFIMA) model. For the purpose of testing robustness, we consider alternative oil price proxies, multiple data samples, and multiple forecast horizons. Our results show impressive performance of macroeconomic variables (output, inflation, and interest rate) in predicting aggregate and sector-level stocks in the US.

The remainder of this paper is structured as follows. Section 2 provides the theoretical background for further justification of the newly introduced variables. Section 3 provides the model set-up, including the forecast performance measures. Section 4 deals with data and preliminary analysis, while empirical results are presented in Section 5. In Section 6, we offer some policy implications of the findings, and Section 7 concludes the paper.

2 Theoretical basis for the predictive model of the stock market

The literature involving macroeconomic analysis of stock returns has relied on APT, which incorporates the role of macroeconomic risks (also described as systematic risks) in stock returns. Other asset pricing models include, but are not limited to, the capital asset pricing model (CAPM) and its variants, particularly those developed by Fama and French (1993, 1995, 1996, 2015, 2017). However, our choice of APT is justified by its ability to estimate the effect of various macroeconomic and financial risks on the stock market; this contrasts with the CAPM, which only recognizes market risk, and with Fama and French, who deal with firm-specific characteristics.

This study introduces three macroeconomic variables—namely inflation, a measure of real economic activity, and interest rate—into the oil-based predictive model for stock returns. Theoretically, economic agents tend to increase investment when they expect a higher inflation rate. This implies that when investors expect higher inflation, demand for investment increases, leading to a rise in prices of stocks. The implication of this transmission channel is that inflation expectations can help in predicting current levels of stock prices. Moreover, it has been shown empirically that, among other macroeconomic variables, the inflation rate tends to have the highest predictive power in the forecast model for the stock market (see Liu et al., 2015). The second

5

macro variable introduced into our oil-based model for stock returns is a measure of real economic activity, with the industrial production index as a proxy. The theoretical relationship between the stock market and real economic activity can be explained by demand-side shock, which assumes that a positive productivity shock caused by positive demand shock, for instance, leads to higher revenues and profits, which in turn causes an increase in dividends and, therefore, in stock prices (see Tsagkanos and Siriopoulos, 2015). Hence, real economic activity may be expected to predict the stock market. The third macro variable considered is interest rate, with short-term interest rate as a proxy. An increase in interest rates, perhaps occasioned by the need to control money supply and, by extension, inflation, may cause earnings to fall and stock prices to drop, while a fall in interest rates may encourage increased spending and cause stock prices to rise.

3 The model and estimation procedure

3.1 Augmented linear and nonlinear covariate models

Following the literature, the conventional bivariate single predictive regression model in equation (1) postulates oil price as the main predictor of the US stock market:

$$r_t = \alpha + \lambda p_{t-1} + \varepsilon_t \tag{1}$$

where r_t and p_t are logs of stock prices and oil price, using the West Texas Intermediate (WTI) crude oil price, and ε_t is the zero-mean idiosyncratic error term on stock returns. The coefficient λ essentially indicates factor loading on oil prices (Chen et al., 1986; Shanken and Weinstein, 2006). The underlying null hypothesis of no predictability is that $\lambda = 0$. This specification is in line with the approaches of Narayan and Gupta (2015) and Phan et al. (2015) for predicting the US stock market. Note that equation (1) is a linear version of the single-factor predictive model. The underlying estimation procedure that allows for endogeneity, persistence, and conditional heteroscedasticity effects is discussed in the Appendix.

The proliferation of evidence for the influence of the macroeconomic environment on stock price dynamics justifies the inclusion of the three additional macro predictors in our forecast model. Thus, we hypothesize that the inclusion of the macroeconomic factors in equation (1) is likely to enhance the predictability of oil price for stock returns. Therefore, equation (1) is extended to include interest rate, inflation, and the industrial production index. The resulting equation (2) is expressed as follows:

$$r_t = \alpha + \lambda p_{t-1} + \gamma' Z_{t-1} + \varepsilon_t \qquad (2)$$

where $Z_t = \begin{bmatrix} y_t & i_t & \pi \end{bmatrix}'$ and y_t is the log of the US industrial production index, i_t is the log of the interest rate, and π is the log of the consumer price index; $\gamma = 0$ is the null hypothesis of no predictability in the relevant variables. However, some of the extant literature on the oil–stock relationship (including Kilian and Park, 2009; Lee and Chiou, 2011; Chang and Yu, 2013; Huang et al., 2017) has pointed out that a linear specification of the relationship may not be reasonable, since the response of the stock market to changes in oil price can be either positive or negative depending on whether the changes are driven by demand- or supply-side shocks. Hence, it is imperative to accommodate this type of nonlinearity in a predictive regression framework. Consequently, we follow the approach of Shin et al. (2014) to disaggregate the oil price into positive and negative price components, as defined in equations (3) and (4):

$$p_t^+ = \sum_{j=1}^{t} \Delta p_j^+ = \sum_{j=1}^{t} \max\left(\Delta p_j, 0\right) \qquad (3)$$

$$p_t^- = \sum_{j=1}^{t} \Delta p_j^- = \sum_{j=1}^{t} \min\left(\Delta p_j, 0\right) \qquad (4)$$

where p^+ and p^- are defined as positive and negative partial sum decompositions of oil price changes. Relying on the assumption that the multi-factor predictive model in equation (2) will outperform the single-factor model in equation (1), we further examine whether oil price predicts the US stock market more accurately when the former model is formulated in nonlinear form, as in equation (5):

$$r_t = \alpha + \lambda^+ p_{t-1}^+ + \lambda^- p_{t-1}^- + \gamma' Z_{t-1} + \varepsilon_t; \qquad (5).$$

Equation (5) is the nonlinear variant of equation (2). The null hypotheses of no predictability for p_t^+ and p_t^- imply that $\lambda^+ = 0$ and $\lambda^- = 0$, respectively.

In addition, rather than capturing any possible shift in stock returns using the approach of Shin et al. (2014), we allow for two significant structural breaks in equation (2). The breaks are determined by following the GARCH-based unit root test of Narayan and Liu (2015), which allows for two structural breaks. As highlighted in the data section, the choice of the GARCH-based unit root test is underscored by the high-frequency data used, which is known to be susceptible to a conditional heteroscedasticity effect. Thus, in the presence of the GARCH effect, the test is

considered to be more suitable than the conventional unit root tests, such as those of the ADF type. Thus, like equation (3), equation (2) is extended, in equation (6), to account for the structural breaks:

$$r_t = \alpha + \lambda p_{t-1} + \gamma' Z_{t-1} + \psi' D_t + \varepsilon_t \tag{6}$$

where the inclusion of structural shift in the model is denoted by $\lambda' D_t$; $D_t = \begin{bmatrix} D_{1t} & D_{2t} \end{bmatrix}'$ and $D_{it} = 1$ if $t \geq$ the break date and zero otherwise, where $i = 1, 2$. A related study by Salisu and Oloko (2015) showed that ignoring structural breaks has implications on hedging effectiveness, even if the analysis is restricted to in-sample estimation. Theoretically, detecting and accounting for structural breaks if they exist may matter for both in-sample and out-of-sample forecasts, particularly for high-frequency series (for example, Paye and Timmermann, 2006; Rapach and Wohar, 2006, 2008; Pesaran and Timmermann, 2007; Giacomini and Rossi, 2009). Recent developments suggest that accounting for a structural break if it exists, even when modeling with time-varying parameter regressions, makes forecasts more precise (Pesaran and Timmermann, 2007; Pesaran et al. 2013). To verify whether the inclusion of the breaks is valid, we compare the forecast performance of equation (6) with equation (2) for both the in-sample and out-of-sample forecasts.

3.2 Estimation procedure and measures of forecast performance

We begin the analysis by testing for the presence of endogeneity, persistence, and conditional heteroscedasticity effects in the predictors, as these effects are found to have implications for the forecast performance (Westerlund and Narayan, 2015). In the Appendix, we offer theoretical expositions for the implementation of the WN estimator for both single-factor and multi-factor model set-up. Thus, in the presence of any of these effects, the WN estimator is considered appropriate. This evidence has been empirically validated in the works of Westerlund and Narayan (2012, 2015), Narayan and Gupta (2015), Phan et al. (2015), Salisu and Isah (2017), and Salisu et al. (2018), among others. We employ the recursive window approach to generate the forecasts; this approach is preferred to the fixed parameter method, as it also accounts for the time-varying nature of the estimates. This is particularly important when dealing with long time series that may have witnessed structural changes. The recursive window approach is an expanded window approach that adds new observation(s) to the sample and re-estimates thereafter. This process is followed continuously until the end of the forecast period is reached.

For forecast evaluation, two forecast measures are used: the Campbell and Thompson (2008) test (hereafter, C-T test) and the Clark and West (2007) test (hereafter, C-W test). The C-T test, described as the out-of-sample $R^2 (OOS\_R)$ statistic, is computed as $OOS\_R = 1 - (M\widehat{SE}_1 / M\widehat{SE}_0)$, where $M\widehat{SE}_1$ and $M\widehat{SE}_0$ are the mean square error (MSE) of the out-of-sample prediction from the unrestricted and restricted models, respectively. Hence, in the context of this study, the restricted model consists of the single-factor variants, while the unrestricted version consists of the multiple-factor models. A positive value of the statistic (i.e., $OOS\_R > 0$) suggests that the unrestricted model outperforms the restricted model; otherwise, it does not. The C-W test, which compares the predictive accuracy of two competing models, is also employed. Unlike the C-T test, the C-W test is able to determine whether the difference between the forecast errors of the two models is statistically significant or different from zero. The underlying procedure for the C-W test involves calculating equation (7):

$$\hat{f}_{t+k} = \left(r_{t+k} - \hat{r}_{1t,t+k} \right)^2 - \left[\left(r_{t+k} - \hat{r}_{2t,t+k} \right)^2 - \left(\hat{r}_{1t,t+k} - \hat{r}_{2t,t+k} \right)^2 \right] \qquad (7)$$

where k is the forecast period, $\left(r_{t+k} - \hat{r}_{1t,t+k} \right)^2$ is the squared error for the restricted model (i.e., model 1), $\left(r_{t+k} - \hat{r}_{2t,t+k} \right)^2$ is the squared error for the unrestricted model (model 2), and $\left(\hat{r}_{1t,t+k} - \hat{r}_{2t,t+k} \right)^2$ is the adjusted squared error introduced by the C-W test to correct for any noise associated with the larger model's forecast. Thus, the sample average of $\hat{f}_{t+k}$ can be expressed as $MSE_1 - (MSE_2 - \text{adj.})$, and each term is computed as

$$MSE_1 = P^{-1} \sum \left(r_{t+k} - \hat{r}_{1t,t+k} \right)^2,$$

$$MSE_2 = P^{-1} \sum \left(r_{t+k} - \hat{r}_{2t,t+k} \right)^2, \text{ and}$$

$$\text{adj.} = P^{-1} \sum \left(\hat{r}_{1t,t+k} - \hat{r}_{2t,t+k} \right)^2,$$

where P is the number of predictions used in computing these averages. To test for equality of forecast performance between models 1 and 2, the $\hat{f}_{t+k}$ is regressed on a constant, and the resulting t-statistic for a zero coefficient is used to draw the inference. Since the null hypothesis tests for equality of MSEs, the alternative hypothesis implies otherwise. The null hypothesis is rejected if

the test statistic is greater than +1.282 (for a one-sided 0.10 test) or +1.645 (for a one-sided 0.05 test).

As part of our model validation procedure, the forecast results of all the economic models considered as previously described are compared with the ARFIMA model, which seems to be representative of univariate time series models and the historical average model that is a popular benchmark in the forecasting literature. The validation is carried out for both in-sample and out-of-sample forecasts. For robustness, we consider multiple sample periods, multiple forecast horizons, and multiple forecast measures. For the multiple sample periods, we consider 50% and 75% of total observations, and the analyses are carried out singly for these data samples. The use of multiple in-sample periods has become standard practice for evaluating the forecast performance of predictive regression models. The out-of-sample forecast performance is evaluated on the basis of three forecast horizons: 10, 20, and 30 months ahead. Two benchmark crude oil prices, WTI and Brent, are used as alternative oil price proxies, the former for the main analyses and the latter for robustness checks.

4 Data and preliminary analyses

4.1 Data issues

To make a baseline case of the single-factor predictive model proposed by Narayan and Gupta (2015), the US S&P 500 index and WTI oil price are used. The Shin et al. (2014) method is used to decompose the oil price into positive and negative oil price. Hence, this study makes use of monthly data for WTI oil price and US S&P 500 index (large cap stocks).[5] As we extend our model to analyze sectoral stock predictability, the stock performance of 10 sectors with high stock market turnovers are considered. The relevant sectors are: consumer discretionary, consumer staples, energy, finance, health care, industries, information technology, materials, telecommunication services, and utilities. The macroeconomic variables, whose predictive powers on the aggregate and sectoral stocks are examined, are output (represented by IPI, the US Industrial Production Index), inflation rate (proxied by CPI, the US Consumer Price Index), and interest rate (with the three-month Treasury Bill Rate as proxy).

---

[5] We are being specific about the categorization of stocks used, as large cap and stock cap stocks are not necessarily expected to react in a similar way to oil price shocks (see Salisu et al., 2017).

Table 1

Data coverage.

Items	In-sample period				Out-of-sample period					
	Start date: / End date	Obs	50% End date: Obs	75% End date: Obs	50% End date: Obs			75% End date: Obs		
					h = 1	h = 3	h = 6	h = 1	h = 3	h = 6
Aggregate	1947M01:2017M03	843	1982M02:422	1999M09:633	1982M03:423	1982M05:425	1982M08:428	1999M10:634	1999M12:636	2000M03:639
SECTORS										
Cons_disc	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Cons_staples	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Energy	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Finance	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Health	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Industries	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Info_tech	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Materials	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Telecom	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251
Utility	1990M01:2017M03	327	2003M08:164	2010M05:245	2003M09:165	2003M12:167	2004M03:170	2010M06:246	2010M08:248	2010M11:251

The data for the aggregate stock (US S&P 500) and the sectoral stocks were obtained from the Bloomberg Terminal, and the data for WTI oil price were sourced from the Federal Reserve Bank of St. Louis (FRED). The periods of analysis for the aggregate and sectoral stocks differ, owing to differences in historical availability of aggregate and sectoral data for US stocks. Specifically, the analysis for aggregate stocks covers the period January 1947 to March 2017, while that of the sectoral stocks covers January 1990 to March 2017. Thus, we have 843 and 327 observations for aggregate and sectoral stocks, respectively, each of which is large enough to allow reasonable inferences to be drawn.

Meanwhile, as the main objective of this study is to determine an appropriate model for predicting US stock returns, we follow the standard approach for predictive model analysis by dividing our full sample period into 50% and 75% for in-sample analyses (see also Westerlund and Narayan, 2014). For out-of-sample analysis, however, we consider forecasts at one month (h = 1), three months (h = 3) and six months (h = 6) ahead. The three-interval forecast horizons were chosen to test the long-term persistence of the out-of-sample predictive capacity of the model. The summary of data coverage for aggregate and sectoral stocks under full sample, in-sample, and out-of-sample forecast analysis is presented in Table 1.

Figure A1 (in the Appendix) gives a graphical representation of the data employed in the analysis. The 11 rows and four columns detail the trend in the relationship between US stock prices and the four predictors. The first row of the figure presents the graphs for the aggregate stock market in relation to oil price and other predictor variables, while the subsequent 10 rows are the corresponding graphs for each sector. The relationship between stocks and WTI oil price, economic output (IPI), inflation rate (CPI), and interest rate (INT) are presented in the first, second, third, and fourth columns, respectively. For asymmetric analysis, oil price was decomposed into positive and negative changes, following the approach of Shin et al. (2014).

Figure A1 indicates inconsistency in the relationship between aggregate US stock and oil price; it is sometimes positive and sometimes negative. Relevant analysis of the periods of positive and negative oil–US stock trends can be found in Salisu and Oloko (2015). The inconsistency in the relationship may be due to the dual nature of oil as import and export commodity in the US economy, which gives oil shocks the potential to affect aggregate demand and supply. For instance, an increase in shale oil production coupled with a reduction in US imports of crude oil has been proposed as one of the main causes of the 2014 oil price crash (Baffes et al., 2015). Therefore, the

relationship may be positive when a booming US economy causes increased demand for oil, leading to higher oil prices; it may be negative if high oil prices depress the US economy through high production costs. In particular, the issue of endogeneity in the oil–US stock relationship has been recognized and addressed using appropriate methodology.

While oil price shows an inconsistent trend in relation to stock prices, output level and inflation rate show a clear positive trend relationship, suggesting that they are closely related with stock prices and may improve the forecast performance of stock market predictive model. The positive relationship is not unexpected, as the stock market is a good indicator of economic productivity (Solnik, 1987), and the inflation rate increases general price levels, including stock prices. The relationship between oil price and sectoral stock prices is inconsistent, but some sectors appear to exhibit stronger associations than others. For instance, the positive association of oil price with energy, finance, and materials stocks (rows 4, 5, and 9 of Figure A1) appears stronger than the association of oil price with stocks of any other sector. This may partially confirm the findings of Yu et al. (2018) that energy sector stocks are the most responsive to oil price changes and the most prone to oil price risk.

The nature of the trend association of output and inflation rate with stock prices at sectoral level appears to be sector-specific. Specifically, while the stock prices of almost all sectors move on average in the same direction as stock prices (column 2 of Figure A1), the stock prices of some sectors do not seem to respond significantly to the rising trend in price level. Relevant examples include the finance, information technology, telecommunications, and utility sectors (rows 5, 8, 10, and 11). The relationship between interest rate and stock prices (the last column) appears negative both on aggregate and at the sectoral level. The negative trend in the relation between stock prices and interest rate (three-month Treasury Bill Rate) can be explained by US investors using the two investment types more as substitutes than as complements. However, as trends do not ordinarily explain causality, the real relationship between stocks and macroeconomic variables in-sample and out-of-sample is analyzed in the subsequent sections.

4.2 Descriptive statistics

We conduct descriptive statistics to describe the moments and statistical features of our data. Usually, for high-frequency time series, stylized facts suggest the inherent presence of conditional heteroscedasticity, or high volatility, and fat-tailedness, or nonnormality (see Enders, 2014). In addition, Westerlund and Narayan (2015) show that predictability models of oil–US

stock relationships suffer from the problem of endogeneity. The essence of the current section and the preliminary analysis section that follows is to confirm some of these features in our data. Table 2 presents descriptive statistics of selected variables used in our aggregate and sectoral predictive models.

Table 2

Descriptive statistics.

	Mean	Max.	Min.	Std. Dev.	Jarque–Bera	CV
US STOCK						
Aggregate (S&P 500)	494.29	2363.64	14.00	592.91	210.81***	119.95
Cons_disc	258.48	700.24	53.75	154.57	71.71***	59.80
Cons_staples	246.90	565.79	55.44	127.22	29.82***	51.53
Energy	312.90	727.63	83.35	185.96	30.27***	59.43
Finance	154.63	266.70	87.41	43.32	18.19***	28.02
Health	350.63	884.71	65.86	205.85	40.35***	58.71
Industries	252.04	563.99	61.03	120.79	11.08***	47.92
Info_tech	349.96	916.71	56.96	213.16	20.59***	60.91
Materials	172.00	328.73	59.50	70.79	23.71***	41.16
Telecom	143.03	330.62	77.28	51.81	247.95***	36.22
Utility	154.63	266.70	87.41	43.32	18.19***	28.02
West Texas Intermediate (WTI) oil price						
Aggregate	24.31	139.96	1.62	27.52	508.67***	113.20
Sectors	46.91	139.96	11.37	30.59	39.09***	65.21
US Industrial Production Index (IPI)						
Aggregate	57.33	106.61	13.63	29.71	65.52***	51.82
Sectors	89.83	106.61	62.43	13.51	40.38	15.04
US Consumer Price Index (CPI)						
Aggregate	104.57	243.80	21.50	73.94	83.53***	70.71
Sectors	187.72	243.80	127.40	34.67	23.87***	18.47
US three-month Treasury Bill Rate (INT)						
Aggregate	4.13	16.30	0.01	3.10	161.28***	75.06
Sectors	2.80	7.90	0.01	2.30	24.73***	82.14

Note: CV denotes coefficient of variation. *** indicates significance at the 1% level.

The figures for the Jarque–Bera statistic in Table 2 show that the aggregate and sectoral indexes are nonnormal, with the exception of industrial production. The nonnormality observed in these series is typical of long, high-frequency financial and economic time series. The results show the average value of the S&P 500 to be 494.29 basis points, with a standard deviation of 592.91 and the highest variability (shown by coefficient of variation (CV) of 119.95) in the period under consideration. The figures in the last column show that the second highest variability (after the S&P 500 aggregate index) was in WTI oil price, and the three-month interest rate was third highest. At the sectoral level, the stocks for the IT sector (CV = 60.91) show the highest variability, followed by consumer discretionary (CV = 59.80) and consumer staples (CV = 59.43).

With the exception of a few sectors, such as finance, energy, and telecommunications, most sectors performed at their historical best, as shown in the last month of the analysis (see Figure A1). Evidently, virtually all the series are nonnormal, as expected for most high-frequency financial and economic series. Thus, in the determination of the weight used in transforming the original predictive model, we consider a nonnormal error distribution that mimics the distribution of the regression error in order to produce more precise and accurate forecasts.

4.3 Preliminary analyses

We carry out preliminary tests for the presence of significant autocorrelation, heteroscedasticity, persistence, and endogeneity effects in the predictors of aggregate and sectoral US stocks. All variables are expressed in natural logarithms for ease of interpretation, and oil price is decomposed into negative and positive oil price to account for nonlinearity in oil price effects (Westerlund and Narayan, 2015). The formulation for decomposing oil price changes is specified in equations (3) and (4), following Shin et al. (2014). The results of the test for ARCH effect, (autocorrelation, persistence, and unit root) are presented in Table 3, and the results for endogeneity tests are presented in Table 4.

Table 3

Results for ARCH effect, autocorrelation, persistence, and unit root tests.

Items	ARCH(5)	ARCH(10)	LB(5)	LB(10)	Persistence	Unit root
LSTOCK						
Aggregate	6.1961***	4.0209***	10.495*	14.475		-5.136***[b]
Cons_disc	5.8369***	3.5589***	4.5633	6.9213		-5.247***[b]
Cons_staples	2.6634**	1.2435	5.4828	9.5401		-4.007**[b]
Energy	2.1271*	1.9126**	1.7402	4.0312		-4.161**[b]
Finance	8.1820***	4.7795***	3.3963	11.840		-4.005**[b]
Health	2.6930**	1.4275	9.8384*	23.363***		-4.302**[b]
Industries	10.985***	5.8853***	3.6549	13.197		-4.125**[b]
Info_tech	20.846***	12.375***	4.3115	8.3679		-3.965**[b]
Materials	8.4071***	4.9705***	0.9079	10.680		-6.515***[b]
Telecom	9.7046***	8.1588***	5.2236	14.208		-4.873***[b]
Utility	8.1820***	4.7795***	3.3963	11.840		-4.604***[b]
LWTI						
Aggregate	2.7995***	1.5322	23.275***	29.413***	0.9973***	-7.933***[b]
Sectors	6.7582***	3.2698***	20.072***	25.893***	0.9888***	-5.155***[b]
LIPI						
Aggregate	25.165***	12.565***	231.17***	237.27***	0.9983***	-4.356***[a]
Sectors	14.670***	8.0593***	117.66***	139.84***	0.9943***	-4.565***[a]
LCPI						
Aggregate	40.807***	28.060***	450.87***	702.68***	0.9997***	7.544***[b]
Sectors	18.488***	9.3231***	92.483***	130.74***	0.9972***	-3.869***[b]
LINT						
Aggregate	64.298***	32.692***	23.275***	29.413***	0.9912***	-8.0353***[a]
Sectors	11.237***	22.846***	27.089***	29.583***	0.9873***	-4.122***[a]

Note: The values reported are the F-statistics for the ARCH effect test and Ljung–Box Q statistics for the autocorrelation test. We consider two different lag lengths (k) of 5 and 10 for robustness. The null hypothesis for the autocorrelation test is that there is no serial correlation, while the null for the ARCH test is that there is no conditional heteroscedasticity. The persistence effect is captured by the coefficient of the first-order autoregressive process for the predictors. Asterisks ***, **, and * indicate rejection of the null hypothesis at significance levels of 1%, 5%, and 10%, respectively. The unit root test follows the GARCH-based unit root test developed by Narayan and Liu (2015)

with the underlying test equation $\Delta y_t = \lambda_0 + \lambda_1 t + \delta y_{t-1} + \sum_{i=1}^{k} D_i B_{it} + \varepsilon_t;$, where y_t denotes the series under

consideration, t is a time trend, $B_{it} = 1$ if $t \geq T_{B_i}$, and $B_{it} = 0$ otherwise. The null hypothesis of unit root given as

$H_0: \delta = 0$ is tested against the alternative hypothesis of stationarity denoted as $H_1: \delta < 0$. Superscript letters [a] and [b] denote $I(0)$ and $I(1)$, respectively.

Table 3 reveals that virtually all the predictors suffer an ARCH effect and autocorrelation. The null hypothesis of no ARCH effect is strongly rejected at conventional levels of significance for oil price and macroeconomic variable predictors of aggregate and sectoral US stocks. Similarly, the null hypothesis of no serial correlation by Ljung–Box statistics is strongly rejected at conventional levels of significance. This suggests that accounting for ARCH effect and autocorrelation in modeling the relationship between aggregate and sectoral stock prices and macroeconomic variables is important for the avoidance of biased results. Hence, it is important to choose an estimator consistent with autocorrelation and conditional heteroscedasticity.

For the unit root test, we employ the GARCH-based unit test of Narayan and Liu (2015), which is more appropriate for the series we are dealing with. One of the shortcomings of the conventional unit root tests, particularly the ADF type, is their inability to exploit the information contained in conditional heteroscedasticity in the test equation. This is one of the advantages of the GARCH-based unit root tests; they allow the error from the test equation to follow a GARCH process and are thus able to account for conditional heteroscedasticity in the estimation process. There are different versions of this type of unit root test, such as those developed by Cook (2008), Narayan and Liu (2011, 2013, 2015), and Narayan et al. (2016). However, Narayan and Liu's 2015 version is preferred to the others, as it allows for both structural breaks and time trend in the test equation. While Cook (2008) ignores both features, the tests of Narayan and Liu (2011, 2013) and Narayan et al. (2016) allow for structural breaks only.[6] A similar test with structural breaks and time trend was proposed by Narayan and Popp (2010); however, it ignores conditional heteroscedasticity and may not be suitable for high-frequency series such as those used in our paper.

The results of the GARCH-based unit root test are presented in Table 3. With the exception of interest rate and output, which can be described as $I(0)$, all the other series are $I(1)$. This outcome further validates the need to perform a separate test for persistence effect. One may conclude from the unit root test that interest rate and output do not suffer severe persistence effects; however, in terms of magnitudes based on the first-order autoregressive regression, the effects are not as small as suggested by the unit root test. In other words, a series being $I(0)$ does not imply zero persistence effect, and an additional test is required to determine the extent of the effect.

---

[6] See Salisu and Adeleke (2016) and Salisu et al. (2016) for recent applications of the GARCH-based unit root tests.

Table 4

Results for endogeneity test.

LSTOCK	LWTI	LWTI_P	LWTI_N	LIPI	LCPI	LINT
Aggregate	0.4058	1.0239***	−1.8539***	4.0519***	−21.6388***	−0.3732
Cons_disc	0.3607	0.1827	0.1566	6.9861***	3.2684	0.0323
Cons_staples	0.2108	−0.0054	−0.0970	3.0935*	−2.1514	−0.0482
Energy	0.5805***	0.3931**	0.1413	−2.2348	11.050***	−0.1459
Finance	0.0250	−0.0617	−0.1744	−1.3375	2.4352	−0.0140
Health	0.3046	0.3888	−0.2611	1.5418	0.1487	−0.0294
Industries	0.3702	0.2763	0.1038	4.7742***	4.4695	−0.0132
Info_tech	0.5698	0.9320**	−0.0052	8.0793***	5.2637	0.0183
Materials	0.3537**	−0.0021	0.3733**	5.1958***	6.9918***	−0.0191
Telecom	0.1337	0.4415	−0.0326	5.1807**	0.3473	0.0194
Utility	0.0250	−0.0617	−0.1743	−1.3375	2.4352	−0.0140

Note: The endogeneity test follows a three-step procedure. First, we run the following predictive regression model: $r_t = \alpha + \beta x_{t-1} + \varepsilon_{r,t}$, where r_t represents stock returns, and x_{t-1} is the predictor variable, such as oil price and its asymmetry, output, inflation, and interest rate. In the second step, we follow Westerlund and Narayan (2015) and model the predictor variable as follows: $x_t = \mu(1-\rho) + \rho x_{t-1} + \varepsilon_{\pi,t.}$. In the final step, the relationship between the error terms is examined using the following regression: $\varepsilon_{r,t} = \phi \varepsilon_{x,t} + \eta_t$. The reported values are the coefficients of ϕ. If this is found to be statistically significant and different from zero at any of the conventional levels of significance, the predictor is endogenous. Asterisks ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

The results of the persistence test in Table 3 show that the predictors have high persistence, as they are very close to unity. Table 4 makes it clear that there are endogeneity issues, with few predictors that need to be captured in the model. For the predictors of US stocks at aggregate level, negative and positive oil price, output, and inflation are endogenous. This indicates the possibility of reverse causality between stock prices and these predictors. Interest rate is not endogenous to US stock prices at aggregate or sectoral levels. However, endogeneity is evidence for oil price and asymmetric oil price with stocks of energy and material sectors. Output is endogenous with stocks for six out of the 10 sectors, and inflation is endogenous with stocks for the energy and materials sectors.

This suggests that an estimator consistent with endogeneity must be chosen to avoid endogeneity bias in the empirical results. The competing estimators are Lewellen (2004) and

Westerlund and Narayan (2012); however, to further account for heteroscedasticity in the series as noted previously, Westerlund and Narayan (2012) is favored.

5 Discussion of results

Following our preliminary analyses, which confirm the presence of persistence and conditional heteroscedasticity in the US aggregate and sectoral stock returns and, to some extent, endogeneity between these stock returns and the proposed predictors, we favor the use of the WN estimator. As stated earlier, this estimator is better that the Lewellen (2004) alternative, which does not capture the presence of ARCH effects (Westerlund and Narayan, 2015). Basically, we estimate seven different models: five for the empirical analysis and two others for model validation and robustness. The first model (denoted as k_S) is the oil price-based single-factor predictive model. This has been identified as a very good model for predicting US stock returns, albeit at the aggregate level (see Westerlund and Narayan, 2015).

As this study examines the ability of macroeconomic factors to predict US stock returns using the single-factor predictive model as a base, it would empirically validate whether accounting for the role of macroeconomic variables in an oil price-based single-factor predictive model would improve the forecast performance of the model or lead to its over-fitness. Hence, the second model, which is the main concern of this study, is the multiple-factor predictive model (denoted as k_M). This is compared with the first model (k_S) to examine whether accounting for macroeconomic factors improves or degrades the predictability of an oil price-based single-factor predictive model for US aggregate and sectoral stock returns.

The third model (k_B) investigates whether the forecast performance of our multiple-factor predictive model could be improved when structural break is accounted for in the model. This involves introducing additional variables in the form of dummy variables to capture the effect of structural breaks. The presence and significance of structural breaks in the aggregate and the sectoral US stock returns was determined using the Bai and Perron (2003) multiple breakpoints test. This model will be compared with k_M to determine the significance of structural breaks in a multiple predictive model for US stock returns.

The fourth and fifth models, like the third model, seek to examine probable improvement in the predictability of our multiple-factor model by accounting for some special data properties. Specifically, the fourth and fifth models examine the potential for obtaining a better multi-factor

predictive model by accounting for the role of oil price asymmetry. According to Westerlund and Narayan (2015), in their single-factor predictive model, oil price was found to predict US stock returns asymmetrically, with negative changes in oil price predicting better than positive changes. In the context of this study, the fourth model (k_P) examines the predictability of positive oil price changes, while the fifth model (k_N) investigates the predictability of negative oil price changes. Both would be compared with the multiple-factor model to examine the relevance of accounting for oil price asymmetry in predicting US stock returns. The two other models for model validation and robustness are the statistical model following ARFIMA (k_{ARF}) and the historical average model (k_{AVR}).

For all the models, the in-sample analyses are performed with 50% and 75% of the data sample, and out-of-sample analyses are performed with forecasts one month, three months, and six months ahead. The Root Mean Square Error (RMSE), which is the usual means of evaluating the forecast performance of predictive models, was employed. However, as this method is incapable of assessing the significance of the difference between two competing predictive models, an alternative forecast evaluation method proposed by Clark and West (2007), the C-W test, is employed. This method is most appropriate for dealing with nested models (as in this study), particularly as it adjusts for the noise associated with the forecasts of larger models, unlike the alternative test proposed by Diebold and Mariano (2002).

Thus, the presentation and discussion of the results of this study begins with the analysis of the in-sample forecast performance of our empirical models. This is followed by the analysis of the out-of-sample forecast performance and then the analyses of model validation and robustness.

5.1 In-sample forecast evaluation for aggregate and sectoral US stock returns

For the in-sample forecast performance evaluation of our empirical predictive models, we present the RMSE of the respective models in Table 5. Evidently, the RMSE explains the deviation of the forecast series from the actual. Zero RMSE will imply a perfect forecast; hence, the closer the RMSE to zero, the better. This presupposes that a model with lower RMSE is better than a model with higher RMSE. The results show that the RMSE for aggregate stock returns in the single-factor predictive model (k_S) is higher than that of the multiple-factor model (k_M) and its extensions under the 50% sample size. This indicates that the multiple-factor model is better, even if the extension of the multiple-factor model to account for structural break and asymmetry does

not seem to improve its forecast performance. A similar conclusion can be drawn from the 75% sample size. The RMSE for k_S is higher than that for k_M, which indicates that k_M is better than k_S. However, the extended multiple-factor models produced a higher RMSE than that of the parsimonious multiple-factor model, implying that structural breaks and asymmetry constitute noise to the multiple predictive model.

Table 5

In-sample RMSE for predictive models.

Stock returns	50% sample size					75% sample size				
	k_S	k_M	k_B	k_P	k_N	k_S	k_M	k_B	k_P	k_N
Aggregate	**0.8274**	**0.2526**	0.2547	0.2381	0.3841	**0.5148**	**0.3960**	0.5368	0.5021	0.5474
SECTORS										
Cons_disc	**0.1747**	**0.1531**	0.1892	0.2158	0.3182	0.3002	0.4836	0.4253	0.3537	0.4857
Cons_staples	0.2344	0.2950	0.2951	0.1699	0.1005	**0.3697**	**0.2771**	0.2710	0.2463	0.2386
Energy	0.1512	0.1543	0.1543	0.2376	0.2082	0.1666	0.3016	0.2999	0.2231	0.2920
Finance	0.1845	0.2051	0.1021	0.1766	0.1693	0.3893	0.4274	0.3738	0.5299	0.5406
Health	**0.3576**	**0.1951**	0.1951	0.2075	0.1999	**0.7098**	**0.3572**	0.5065	0.3394	0.3316
Industries	0.1830	0.1914	0.1677	0.2244	0.2752	0.3617	0.6829	0.5508	0.5998	0.6364
Info_tech	**0.4536**	**0.4153**	0.2608	0.4827	0.5217	0.6044	0.6538	0.6538	0.5086	0.4602
Materials	**0.1635**	**0.1445**	0.1445	0.1259	0.1383	**0.3656**	**0.4280**	0.3899	0.3344	0.3035
Telecom	**0.3395**	**0.3056**	0.2161	0.3291	0.3415	0.4277	0.7390	0.2174	0.7035	0.6773
Utility	0.1845	0.2051	0.1021	0.1766	0.1693	0.3893	0.4274	0.3738	0.5299	0.5406

Note: The reported values are for RMSE from single-factor predictive model k_S, multiple-factor predictive model k_M, multiple-factor model with structural break k_B, multiple-factor model with positive oil price changes k_P, and multiple-factor model with negative oil price changes k_N. Zero RMSE indicates perfect predictability; hence, the closer the RMSE to zero, the better.

On the sectoral stock returns, the RMSE results for the 50% sample size suggest that the multiple-factor predictive model outperforms the single-factor model when predicting stock returns in the consumer discretionary, health, information technology, materials, and telecommunication sectors; for the other sectors—consumer staples, energy, finance, industries, and utilities—the single-factor predictive model appears to perform better. However, as the 75% sample size extends the data coverage to 2010, it appears that the ability of macroeconomic factors to predict sectoral US stock returns was undermined by the occurrence of the global financial crisis. This is indicated, as the stock returns for the consumer discretionary, information

technology, and telecommunication sectors, which were better predicted with the multiple-factor model and the 50% sample size, are now better predicted with the single-factor model and the 75% sample size.

However, as no cognizance is given to the noise usually associated with the larger model's forecast when evaluating the performance of predictive models with RMSE, the conclusion from RMSE may not be totally correct. This prompts the use of the C-W test, which adjusts for the noise associated with the larger model's forecast and assesses the significance of the difference in the forecasting accuracy of two competing models, the parsimonious and the extended or larger model. Based on the initial proposition to determine the better predictive model between single-factor (model 1) and multiple-factor (model 2), between multiple-factor (model 2) and multiple-factor with structural break (model 3), and between multiple-factor (model 2) and multiple-factor with asymmetry (models 4 and 5), identification of the parsimonious and larger models is necessary.

Four cases of model comparison emerged. The first is the case of model 1 vs. model 2 (denoted as $k_M : k_S$). Here, the model for the single-factor case (k_S) is the parsimonious model, having been found to be a good predictive model for US stock returns by Westerlund and Narayan (2015), while the model for the multiple-factor case (k_M) is the larger model. Hence, the multiple-factor model is preferred when the C-W test is positive and significant; otherwise, the single-factor model is preferred. Under the second case, model 3 is compared with model 2 ($k_M : k_B$) to assess the relevance of accounting for structural breaks in the multiple-factor model. In this case, k_M is the parsimonious model, and k_B is the extended model. Thus, if the C-W test is positive and significant, accounting for structural breaks improves the predictability of the particular stock returns; otherwise, accounting for structural breaks is irrelevant in the multiple-factor predictive model.

The third and fourth cases assess the significance of accounting for oil price asymmetry, as they compare the multiple-factor model (with symmetric oil price) with the multiple-factor model with positive oil price changes ($k_P : k_M$) and negative oil price changes ($k_N : k_M$), respectively. In these two cases, our proposed multiple-factor model (with symmetric oil price) is the larger model, as it contains both positive and negative oil price changes; the multiple-factor model with positive oil price changes and the multiple-factor model with negative oil price changes are regarded as the parsimonious models. For asymmetry to improve the forecast performance of

our multi-factor predictive asymmetric model, the positive or negative model or both must perform better than our symmetric multiple-factor model. This presupposes that the C-W test is expected to be negative and significant under either or both cases, $k_P : k_M$ and $k_N : k_M$, for asymmetry to matter; otherwise, our proposed multiple-factor model (with symmetric oil price) is preferred.

Table 6

In-sample results for the Clark and West test.

Stock returnstock returns	50% sample size				75% sample size			
	$k_S : k_M$	$k_M : k_B$	$k_P : k_M$	$k_N : k_M$	$k_S : k_M$	$k_M : k_B$	$k_P : k_M$	$k_N : k_M$
Aggregate	1.6558***	0.0032**	−0.0069***	0.1149***	0.6481***	−0.1056***	0.1122***	0.1710***
SECTORS								
Cons_disc	0.0282***	2.38E-05	0.0416***	0.1340***	0.0830***	0.0641***	−0.0822***	0.0087*
Cons_staples	0.1534***	-	−0.0182***	−0.0053	0.3345***	0.0033***	−0.0142***	−0.0171***
Energy	0.0103***	-	0.0507***	0.0270***	−0.0118**	0.0010***	0.0589***	−0.0001
Finance	0.0032	0.0541***	−0.0074***	−0.0066***	−0.0240***	0.0541***	0.1129***	0.1294***
Health	0.2156***	-	0.0054***	0.0053***	1.3040***	−0.0874***	−0.0112***	−0.0156***
Industries	0.0062**	0.0262***	0.0164***	0.0521***	−0.1389***	0.1938***	−0.0990***	−0.0557***
Info_tech	0.1042***	0.1822***	0.0700***	0.1174***	0.6172***	-	−0.1166***	−0.1063***
Materials	0.0288***	-	0.0025	0.0080***	−0.0388***	0.0363***	−0.0556***	−0.0596***
Telecom	0.0440***	0.0706***	0.0398***	0.0389***	−0.1481***	0.8892***	−0.0509***	−0.0832***
Utility	0.0032	0.0541***	−0.0074***	−0.0066***	−0.0240***	0.0541***	0.1129***	0.1294***

Note: $k_S:k_M$, $k_M:k_B$, $k_P:k_M$, and $k_N:k_M$ represent the four different cases earlier discussed. A positive and significant C-W statistic indicates that the multiple-factor model is preferred to the single-factor model in the first case, and the multiple-factor model with structural breaks is preferred to the multiple-factor model without structural breaks in the second case. A positive and significant C-W statistic in the third and fourth cases indicates that the multiple-factor model is better than the multiple-factor model with structural breaks. The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 6 presents the results of the test for in-sample forecast performance of the competing models under different cases. In the first case ($k_S : k_M$), the C-W statistic is positive and strongly significant for the aggregate and for 80% of sectoral US stock returns. This holds particularly for the 50% sample size. As also suggested by the RMSE results, the ability of macroeconomic factors

to predict sectoral US stock returns appears to have been undermined by the occurrence of the global financial crisis, since the number of sectors rejecting the multiple-factor model increased from two under the 50% sample to six under the 75% sample.

As our predictive model follows the global financial crisis, we would be constrained to draw our conclusions from the results for the 75% sample size whenever the results from the 50% and 75% sample sizes are inconsistent. Hence, on the basis of the C-W test, we conclude from our in-sample analysis that the stock returns for the consumer discretionary, consumer staples, health, and information technology sectors are better predicted using multiple-factor predictive models, whereas stock returns for other sectors—finance, energy, industries, materials telecommunication, and utilities—are better predicted with an oil price-based single-factor predictive model. However, for aggregate stock returns (that is, returns on the S&P 500), the C-W statistic remains positive and significant under the 75% sample, as it is under the 50% sample, indicating that the multiple-factor predictive model outperforms the single-factor predictive model overall.

We now consider the second case, which examines whether accounting for structural breaks in a multiple-factor predictive model will improve its predictive ability. Since stock returns of some sectors are better predicted with a single-factor model, the question of whether accounting for structural breaks in macroeconomic variables augments the predictive model will only apply to the stock returns of the aggregate and the relevant four sectors. The in-sample C-W results in Table 6 show that, for three of the four sectors supporting a multiple-factor predictive model, accounting for structural breaks would improve their forecast performance under both the 50% and 75% samples. Specifically, the predictability of stock returns for the consumer discretionary, consumer staples, and information technology sectors improves when structural breaks are accounted for in their respective multiple-factor predictive models. However, for aggregate stock returns and the health sector, accounting for structural breaks in the multiple-factor model tends to generate noise in the model, as it reduces its fitness.

The relevance of accounting for oil price asymmetry in a multiple-factor predictive model was examined by comparing the multiple-factor model with positive and negative oil price changes with our proposed symmetric multiple-factor model. As in the third and fourth cases, the relevant stock returns are those that perform well in the multiple-factor model; that is, the aggregate (S&P 500) and the consumer discretionary, consumer staples, health, and information technology sectors. As discussed earlier, a positive and significant C-W value implies that asymmetry is not

relevant, while a negative and significant C-W statistic for either or both cases implies that asymmetry matters.

The in-sample results presented in Table 6 show that the C-W statistics for the relevant sectoral stock returns are negative and significant over the longer sample size but show a positive and significant C-W statistic for aggregate stock returns. This indicates that while the stock returns of the four relevant sectors could be adequately predicted with a multiple-factor model augmented with macroeconomic variables, the forecast performance of the stock returns could be improved by capturing the asymmetric effect of oil price changes in the model. For the aggregate stock returns, it suggests that accounting for oil price asymmetric matter in an oil-based single-factor model, as analyzed by Westerlund and Narayan (2015), does not automatically imply that it will matter when using a macroeconomic variable-augmented multiple-factor model. This may mean that accounting for asymmetry in a multiple-factor model results in over-fitness of the model (Salisu and Isah, 2018).

Overall, our in-sample results revealed that macroeconomic variables (output, inflation, and interest rate) play a significant role in improving the predictability of the US aggregate (S&P 500) stock returns and the stock returns of some sectors. In addition, in-sample results show that the predictability of US aggregate stock returns can be improved by accounting for structural breaks and oil price asymmetry. However, the forecast performance of stock returns for the consumer discretionary, consumer staples, and information technology sectors can be improved by accounting for both structural breaks and oil price asymmetry, whereas accounting for oil price asymmetry only is important for improving forecast performance for health sector stock returns.

5.2 Out-of-sample forecast performance

Since a better in-sample forecast does not automatically guarantee a better out-of-sample forecast, we now consider the out-of-sample forecast performance of the models. As for the in-sample analysis, we compare the forecast performance of the single-factor oil price–US stock predictive model by Narayan and Gupta (2015) with our proposed multiple-factor predictive model. In addition, we examine whether accounting for structural breaks and oil price asymmetry leads to a significant improvement in the forecast performance out-of-sample of our proposed multiple-factor model. The decision criteria for the out-of-sample C-W tests remain as discussed under the in-sample analysis. The models are again compared under four cases, namely; model 1 vs. model 2 ($k_M : k_S$), model 2 vs. model 3 ($k_M : k_B$), model 4 vs. model 2 ($k_P : k_M$), and model 5

vs. model 2 ($k_N : k_M$). We use the rolling window approach to conduct forecasts. Additionally, for robustness, we forecast using 50% and 75% of the available data and for forecast horizons one month (h = 1), three months (h = 3), six months (h = 6), and 12 months (h = 12) ahead. This approach is used to evaluate the forecast performance of different predictive models. The analyses of the out-of-sample assessment of the forecast performance of the predictive models are presented in the following sub-sections.

5.2.1 Single-predictor vs. strictly multiple-predictor out-of-sample forecast

Under the in-sample analyses, we find that the multiple-factor predictive model overly outperforms the single-factor predictive model for aggregate (S&P 500) stock returns and for stock returns of four notable sectors: consumer discretionary, consumer staples, health, and information technology. This result needs to be confirmed out-of-sample, as the main motive of prediction is to make out-of-sample forecasts. Table 7 presents the C-W statistics for the four cases of model comparison under 50% and 75% sample sizes, respectively. The performance of the competing models is compared over forecasts one month, three months, six months, and 12 months ahead.

The C-W statistic is positive and strongly significant for the aggregate and for eight of the 10 sectoral US stocks considered under the 50% sample size. However, as with the in-sample analysis, this result is not consistent for all sectors through the 75% sample size. Given this inconsistency, we cannot emphatically conclude that the multiple-factor predictive model will always predict stock returns for the energy, industries, materials, or telecommunication sectors better than the oil price-based single-factor predictive model. Because the C-W statistics of the aggregate stock returns and the stock returns for the consumer discretionary, consumer staples, health, and information technology sectors remained consistently positive and significant over all the forecast horizons considered, we may confirm that the supremacy of the multiple-factor predictive model over the single-factor predictive model in the in-sample also applies out-of-sample.

Hence, we conclude that the role of macroeconomic variables is important and should be considered when setting up a predictive model for S&P 500 (aggregate) stock returns and for stock returns of certain sectors, namely consumer discretionary, consumer staples, health, and information technology. In contrast, the forecast performance of the stock returns of the finance and utility sectors worsened with the introduction of macroeconomic variables, implying that the

oil price-based single-factor predictive model is sufficient for predicting stock returns for these sectors. In other words, accounting for the role of macroeconomic variables constitutes noise in the oil price-based single-factor predictive model and reduce its fitness for forecasting.

Table 7

Clark and West out-of-sample results for single-factor vs. multiple-factor model ($k_S : k_M$).

Stock returns	50% sample size				75% sample size			
	$h=1$	$h=3$	$h=6$	$h=12$	$h=1$	$h=3$	$h=6$	$h=12$
Aggregate	**1.6532*****	**1.6479*****	**1.6395*****	**1.6252*****	**0.6469*****	**0.6448*****	**0.6415*****	**0.6347*****
SECTORS								
Cons_disc	**0.0287*****	**0.0295*****	**0.0308*****	**0.0360*****	**0.0798*****	**0.0739*****	**0.0670*****	**0.0565*****
Cons_staples	**0.1520*****	**0.1494*****	**0.1460*****	**0.1408*****	**0.3326*****	**0.3290*****	**0.3240*****	**0.3152*****
Energy	0.0102***	0.0103***	0.0103***	0.0108***	−0.0126**	−0.0139**	−0.0154***	−0.0164***
Finance	0.0028	0.0022	0.0016	0.0018	−0.0237***	−0.0230***	−0.0215***	−0.0171***
Health	**0.2143*****	**0.2117*****	**0.2077*****	**0.1991*****	**1.3006*****	**1.2938*****	**1.2847*****	**1.2671*****
Industries	0.0064**	0.0068**	0.0077***	0.0130***	−0.1429***	−0.1504***	−0.1605***	−0.1758***
Info_tech	**0.1022*****	**0.0991*****	**0.0956*****	**0.0931*****	**0.6131*****	**0.6052*****	**0.5951*****	**0.5780*****
Materials	0.0292***	0.0296***	0.0298***	0.0315***	−0.0395***	−0.0409***	−0.0426***	−0.044***
Telecom	0.0437***	0.0433***	0.0438***	0.0525***	−0.1510***	−0.1560***	−0.1621***	−0.1705***
Utility	0.0028	0.0022	0.0016	0.0018	−0.0237***	−0.0230***	−0.0215***	−0.0171***

Note: k_S: k_M represents case I, where the single-factor predictive model is compared with the multiple-factor predictive model. A positive and significant C-W statistic indicates that the multiple-factor model preferred to the single-factor model. The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Bold typeface indicates where the multiple-factor predictive model outperforms the single-factor predictive model consistently under the 50% and 75% data samples.

This result has serious implications for stock market hedging and portfolio management, as it provides relevant information to stock market investors about the future performance of the S&P 500 and sectoral stock returns. It shows that the role of macroeconomic variables should not be neglected in a predictive model for S&P 500 stock returns and for the stock returns of the consumer discretionary, consumer staples, health, and information technology sectors; however, it is unimportant in a predictive model for stock returns for the finance and utility sectors. An

increase in macroeconomic uncertainty would significantly affect stock returns for the consumer discretionary, consumer staples, health, and information technology sectors, but it would have no significant effect on stock returns for the finance and utility sectors.

This suggests that equity investment in the finance and utility sectors is less risky than equity investment in the consumer discretionary, consumer staples, health, and information technology sectors. Hence, to hedge unsystematic risk from macroeconomic uncertainty, it would benefit a risk-averse investor to hold more stocks for sectors in which returns are not significantly predicted by macroeconomic variables, such as the finance and utility sectors.

5.2.2 Multiple-predictor model vs. multiple-predictor model with structural breaks

In the in-sample analysis, it was found that the predictability of stock returns for the consumer discretionary, consumer staples, and information technology sectors improves when structural breaks are accounted for in their respective multiple-factor predictive models. However, accounting for structural breaks reduces the fitness of multiple-factor models for aggregate stock returns and for the health sector. This sub-section seeks to confirm these results out-of-sample. The structural breaks were determined using the multiple breakpoints test by Barron and Perron (2003), which determines breakpoints ranging from zero to two.[7] It should be recalled that the objective is to examine whether accounting for structural breaks improves the forecast performance of our proposed multiple-factor predictive model; thus, the focus is on the aggregate stock returns and the stock returns of the relevant four sectors where the multiple-factor predictive model is already established as a parsimonious model.

The Clark and West out-of-sample results presented in Table 8 reveal that the C-W statistics for aggregate stock returns are positive and significant under the 50% sample but negative and significant under the 75% sample. Thus, the significance of structural breaks in the multiple-factor predictive model for aggregate stock returns is inconsistent, which is in line with the in-sample results. For the health sector, both the multiple-factor predictive model and the multiple-factor model with structural breaks are equal, as the Bai and Perron (2003) test shows that no significant structural break exists under the 50% sample. Under the 75% sample, the C-W test is negative and significant over the forecast horizons from one month to 12 months. This indicates

---

[7] The results for our structural break test are presented in Table A1 (in the Appendix).

that accounting for structural breaks is unnecessary, as it reduces the fitness of the multiple-factor model, which is consistent with the conclusion from the in-sample analysis.

Table 8

Clark and West out-of-sample results for multiple vs. multiple with structural breaks ($k_M : k_B$).

Stock returns	50% sample size				75% sample size			
	h = 1	h = 3	h = 6	h = 12	h = 1	h = 3	h = 6	h = 12
Aggregate	0.0031**	0.0032**	0.0032**	0.0032**	−0.1055***	−0.1054***	−0.1052***	−0.1049***
SECTORS								
Cons_disc	−0.0003	−0.0006	−0.0004	0.0007	**0.0669*****	**0.0726*****	**0.0807*****	**0.0962*****
Cons_staples	-	-	-	-	**0.0034*****	**0.0034*****	**0.0035*****	**0.0037*****
Energy	-	-	-	-	0.0010***	0.0010***	0.0010***	0.0009***
Finance	0.0560***	0.0599***	0.0640***	0.06840***	0.0543***	0.0547***	0.0548***	0.0536***
Health	-	-	-	-	−0.0884***	−0.0903***	−0.0923***	−0.0934***
Industries	0.0264***	0.0266***	0.0265***	0.0256***	0.1959***	0.2002***	0.2064***	0.2177***
Info_tech	**0.1865*****	**0.1937*****	**0.2024*****	**0.2186*****	-	-	-	-
Materials	-	-	-	-	0.0377***	0.0403***	0.0442***	0.0511***
Telecom	0.0725***	0.0757***	0.0788***	0.0809***	0.8954***	0.9065***	0.9209***	0.9442***
Utility	0.0560***	0.0599***	0.0640***	0.0684***	0.0543***	0.0547***	0.0548***	0.0536***

Note: $k_M : k_B$ represents case II, where the multiple-factor predictive model is compared with the multiple-factor model with structural breaks. A positive and significant C-W statistic indicates that the multiple-factor model with structural breaks is preferred to the parsimonious multiple-factor model (without structural breaks). The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The dashes (-) in the table indicate that no structural break exists (see Table A1 in Appendix for details of structural break results). Bold typeface indicates where the multiple-factor model with structural breaks outperforms the parsimonious multiple-factor model.

As indicated in bold typeface in Table 8, the consumer discretionary, consumer staples, and information technology sectors remained supportive of structural breaks, with positive and significant C-W statistics. Notably, the relevance of structural breaks in the out-of-sample forecast for stock returns of the consumer discretionary sector increased over time under the 50% sample , manifest after six months. It is, however, consistently beneficial under the 75% sample. This suggests that the in-sample conclusion that accounting for structural breaks improves the performance of multiple-factor predictive model in forecasting stock returns for the consumer discretionary, consumer staples, and information technology sectors continues to hold out-of-sample.

5.2.3 Multiple-predictor model vs. multiple-predictor model with asymmetry

In the single-factor predictive model of Westerlund and Narayan (2015), accounting for oil price asymmetry was found to improve the forecast performance of the model, with negative oil price changes predicting US stock returns better than positive oil price changes. However, for efficient model parsimony, this may not be automatically assumed for the multiple-factor predictive model. This sub-section assesses the significance of accounting for oil price asymmetry out-of-sample to confirm the in-sample result. Notably, cases III and IV are observed mainly for aggregate stock returns and the stock returns of the four sectors accepting the multiple-factor predictive model. Case III compares the strictly multiple-factor predictive model with the multiple-factor model (with positive oil price changes), while case IV compares the strictly multiple-factor predictive model with the multiple-factor model (with negative oil price changes). The C-W statistics for the two cases are presented in Tables 9 and 10.

Table 9

Clark and West out-of-sample results for multiple with positive asymmetry vs. strict multiple $(k_P : k_M)$.

Stock returns	50% sample size				75% sample size			
	h = 1	h = 3	h = 6	h = 12	h = 1	h = 3	h = 6	h = 12
Aggregate	−0.0069***	−0.0069***	−0.0069***	−0.0068***	0.1120***	0.1117***	0.1112***	0.1100***
SECTORS								
Cons_disc	0.0430***	0.0456***	0.0493***	0.0554***	−0.0839***	−0.0873***	−0.0916***	−0.0983***
Cons_staples	**−0.0186***	**−0.0191***	**−0.0195***	**−0.0190***	**−0.0145***	**−0.0150***	**−0.0156***	**−0.0164***
Energy	0.0503***	0.0496***	0.0487***	0.0470***	0.0580***	0.0568***	0.0563***	0.0615***
Finance	−0.0077***	−0.0083***	−0.0088***	−0.0073***	0.1130***	0.1131***	0.1131***	0.1126***
Health	0.0054***	0.0053***	0.0050***	0.0039***	−0.0115***	−0.0119***	−0.0123***	−0.0123***
Industries	0.0166***	0.0170***	0.0176***	0.0183***	−0.1003***	−0.1027***	−0.1061***	−0.1113***
Info_tech	0.0724***	0.0772***	0.0854***	0.1090***	−0.1177***	−0.1196***	−0.1217***	−0.1241***
Materials	0.0024	0.0023	0.0022	0.0021	−0.0559***	−0.0563***	−0.0567***	−0.0569***
Telecom	0.0397***	0.0397***	0.0397***	0.0389***	−0.0509***	−0.0509***	−0.0509***	−0.0505***
Utility	−0.0077***	−0.0083***	−0.0088***	−0.0073***	0.1130***	0.1131***	0.1131***	0.1126***

Note: $k_P : k_M$ represents case III, where the multiple-factor predictive model (with symmetric oil price) is compared with the multiple-factor model (with positive oil price changes). A positive and significant C-W statistic indicates that the multiple-factor model is preferred to the single-factor model. The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Bold typeface indicates where the multiple-factor predictive model with asymmetric positive oil price asymmetry outperforms the proposed multiple-factor predictive model (with symmetric oil price) consistently under the 50% and 75% data samples.

In the in-sample analysis, the results show that the asymmetric effect of oil price improves the forecast performance of the multiple-factor model for predicting the stock returns of the four relevant sectors (consumer discretionary, consumer staples, health, and information technology) over a relatively big sample (75%) but reduces the fitness of the parsimonious multiple-factor predictive model for aggregate stock returns. This sub-section seeks to confirm this result out-of-sample. The results in Table 9 show that, for stock returns for the consumer staples sector, the multiple-factor model with positive oil price asymmetry consistently outperforms the symmetric multiple-factor predictive model. The C-W values for consumer staples are negative and significant over the 50% and 75% samples and over the one-month through 12-month forecast horizons. The in-sample significance for the relevant sectoral stock returns became apparent over the bigger sample, and this was also the case out-of-sample; the C-W statistics for the consumer discretionary, health, and information technology sectors only become negative and significant under the 75% sample.

As with the in-sample analysis, the multiple-factor predictive model for aggregate stock returns does not appear to improve after accounting for structural breaks. Although the positive asymmetric multi-factor model shows significance under 50%, this does not hold over the bigger sample (see Table 9). The consistent superiority of the symmetric multiple-factor model over the multiple-factor model with negative oil price asymmetry is shown in Table 10; accounting for oil price asymmetries does not improve the forecast performance of the multiple-factor predictive model.

Table 10

Clark and West out-of-sample results for multiple with negative asymmetry vs. strict multiple ($k_N : k_M$).

Stock returns	50% sample size				75% sample size			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Aggregate	0.1146***	0.1140***	0.1133***	0.1119***	0.1710***	0.1709***	0.1707***	0.1703***
SECTORS								
Cons_disc	0.1378***	0.1446***	0.1546***	0.1730***	0.0095**	0.0110**	0.0131***	0.0173***

Cons_staples	−0.0048	−0.0032	0.0009	0.0144**	−0.0172***	−0.0172***	−0.0173***	−0.0171***
Energy	0.0273***	0.0280***	0.0284***	0.0276***	−0.0005	-0.0013	−0.0024	−0.0035*
Finance	−0.0070***	−0.0076***	−0.0080***	−0.0060***	0.1292***	0.1286***	0.1277***	0.1258***
Health	0.0053***	0.0053***	0.0057***	0.0067***	−0.0156***	−0.0158***	−0.0158***	−0.0157***
Industries	0.0530***	0.0545***	0.0564***	0.0591***	−0.0558***	−0.0559***	−0.0557***	−0.0546***
Info_tech	0.1222***	0.1319***	0.1480***	0.1918***	−0.1077***	−0.1102***	−0.1130***	−0.1161***
Materials	0.0079***	0.0077***	0.0076***	0.0073***	−0.0600***	−0.0604***	−0.0609***	−0.0611***
Telecom	0.0398***	0.0418***	0.0448***	0.0501***	−0.0836***	−0.0843***	−0.0851***	−0.0858***
Utility	−0.0070***	−0.0076***	−0.0080***	−0.0060***	0.1292***	0.1286***	0.1277***	0.1258***

Note: $k_N : k_M$ represents case IV, where the multiple-factor predictive model (with symmetric oil price) is compared with the multiple-factor model (with negative oil price changes). A positive and significant C-W statistic indicates that the multiple-factor model is preferred to single-factor model. The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Bold typeface indicates where the multiple-factor predictive model with asymmetric positive oil price asymmetry outperforms the proposed multiple-factor predictive model (with symmetric oil price) consistently under the 50% and 75% data samples.

5.2.4 Multiple-factor predictive model vs. historical average model

The historical average model is the usual benchmark for predictive models. Hence, as our multiple-factor predictive model is shown to outperform the oil-based single-factor predictive model in forecasting US aggregate (S&P 500) stock returns and the returns of some sectors, it is pertinent for robustness purposes to examine whether our proposed multiple-factor predictive model also outperforms the historical average model. The C-W test was again used to compare the forecast performance of the two models. In this case, the historical average model is considered as the parsimonious model, while the proposed multiple-factor model is the larger model. Thus, our proposed multiple-factor model outperforms the historical average model if the C-W statistic is positive and significant. The reverse holds if the C-W statistic is negative and significant.

Table 11

Clark and West in-sample and out-of-sample forecast evaluation for historical average vs. multiple-factor model ($k_{AVR} : k_M$).

Stock returns	50% sample size				75% sample size			
	In-sample	h = 1	h = 3	h = 6	In-sample	h = 1	h = 3	h = 6

Aggregate	**0.4543***	**0.4532***	**0.4511***	**0.4478***	**2.5613***	**2.5821***	**2.6247***	**2.6884***
SECTORS								
Cons_disc	**0.1133***	**0.1125***	**0.1112***	**0.1092***	**0.5834***	**0.5873***	**0.5979***	**0.6269***
Cons_staples	**0.3084***	**0.3064***	**0.3030***	**0.2981***	**0.5000***	**0.5014***	**0.5049***	**0.5264***
Energy	0.0375***	0.0372***	0.0367***	0.0365***	0.7229***	0.7268***	0.7357***	0.7528***
Finance	0.1036***	0.1032***	0.1024***	0.1022***	0.1308***	0.1307***	0.1318***	0.1340***
Health	**0.1773***	**0.1762***	**0.1741***	**0.1707***	**0.7460***	**0.7454***	**0.7442***	**0.7436***
Industries	0.1222***	0.1213***	0.1199***	0.1185***	0.5279***	0.5248***	0.5198***	0.5154***
Info_tech	**0.3845***	**0.3820***	**0.3783***	**0.3732***	**1.0272***	**1.0277***	**1.0293***	**1.0354***
Materials	0.0716***	0.0710***	0.0703***	0.0698***	0.1448***	0.1436***	0.1427***	0.1438***
Telecom	0.4160***	0.4131***	0.4081***	0.4027***	0.5902***	0.5878***	0.5848***	0.5831***
Utility	0.1036***	0.1032***	0.1024***	0.1022***	0.1308***	0.1307***	0.1318***	0.1340***

Note: $k_{AVR}:k_M$ represents case V, where the multiple-factor predictive model is compared with the historical average model. A positive and significant C-W statistic indicates that the multiple-factor model is preferred to the historical average model. The reverse holds if the C-W statistic is negative and significant. Asterisks ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. Bold typeface indicates where the proposed multiple-factor predictive model outperforms the historical average models consistently under the 50% and 75% data samples.

Table 11 shows the Clark and West test results for the comparison of the performance of the historical average and multiple-factor models in-sample and out-of-sample. The focus remains on the aggregate stock returns and on the stock returns of the consumer discretionary, consumer staples, health, and information technology sectors, which accept the multiple-factor predictive model as parsimonious, ahead of the single-factor predictive model. The C-W statistics are positive and significant for the aggregate stock returns and for the stock returns of the other four relevant sectors, in-sample, out-of-sample, and over forecast horizons from one month through six months. This implies that, in respect of aggregate stock returns and the stock returns of the consumer discretionary, consumer staples, health, and information technology sectors, the multiple-factor model not only outperforms the single-factor predictive model, it also outperforms the historical average model. This further confirms the superiority of the multiple-factor over oil-based single-factor predictive models in modeling the forecast performance of US aggregate stock returns and the stock returns of selected sectors.

6 Policy implications of the findings

The ability of a stock market investor to accumulate excess returns from equity investment depends on his or her ability to predict the market efficiently. This study carries out empirical analysis to identify the best predictive model for US stock market returns. It identified an oil price-based single-factor predictive model by Westerlund and Narayan (2015) as a base model and investigated whether US macroeconomic conditions have a significant effect on future stock returns by augmenting the oil price-based single-factor predictive model with macroeconomic variables. Our empirical results generate insight and have serious economic implications for fund managers, brokers, and investment analysts in the US stock market. They also have implications for international portfolio investment, as the US economy is regarded as a linchpin of the global economy.

One of the main findings of this study is that the role of macroeconomic variables cannot be underestimated in a predictive model for S&P 500 stock returns and stock returns for the consumer discretionary, consumer staples, health, and information technology sectors. It may, however, be disregarded in a predictive model of stock returns for the finance and utility sectors, among others. This implies that an increase in macroeconomic uncertainty would significantly affect stock returns for the consumer discretionary, consumer staples, health, and information technology sectors but would have no significant effect on stock returns for the finance and utility sectors. This suggests that equity investment in the finance and utility sectors is less risky than equity investment in the consumer discretionary, consumer staples, health, and information technology sectors. Hence, to hedge unsystematic risk due to macroeconomic uncertainty, portfolio investment risk could be minimized by holding more stocks for sectors, such as finance and utilities, in which returns are not significantly predicted by macroeconomic variables.

For the US government, this study shows that macroeconomic factors are important determinants predictors of US stock market performance; hence, any disturbance in macroeconomic fundamentals such as inflation, output, and interest rate will have persistent effects on the US stock market. Therefore, it is recommended that any outbreak of serious economic challenge, such as high inflation, output volatility, or money market uncertainty, should be addressed exigently to forestall its transmission to the US stock market. The collapse of the US sub-prime market and the consequent transmission to the US stock market in 2008 underscores the seriousness of this possibility.

7 Summary and concluding remarks

This study revisits the oil–stock nexus and proposes a predictive model whereby oil price is combined with selected domestic macroeconomic indicators, such as inflation, interest rate, and the level of real economic activities. It contributes to the extant literature on the predictability of US stock market in three ways. First, extending the single-factor oil-based predictive model for the US stock market developed by Narayan and Gupta (2015), we formulate a multi-predictive oil-based predictive model for the US stock market by examining the role of macroeconomic variables. The underlying estimation procedure follows the approach of Lewellen (2004) and Westerlund and Narayan (2015), which accounts for the inherent characteristics of the predictors. While there are multi-predictive models that employ the same approach in the literature, such as Makin et al. (2014), Salisu and Isah (2018), and Salisu et al. (2018), they are not suitable for forecasting stock returns. Second, the study examines whether accounting for asymmetry or structural breaks would improve the forecast performance of the multiple-factor model. Third, it investigates the predictability of the US stock market at both aggregate and sectoral levels. For robustness checks, the analyses were conducted for multiple sample periods and multiple forecast horizons.

For our empirical analyses of in-sample and out-of-sample forecast evaluation, we rely mainly on the test proposed by Clark and West (2007), which adjusts for forecast bias in larger models. Our results reveal that aggregate (S&P 500) US stock returns and stock returns for selected sectors (consumer discretionary, consumer staples, health, and information technology) are better predicted in a multiple-factor model that accounts for macroeconomic variables comprising oil price, inflation, level of real economic activities, and interest rate. This finding is explained by the fact that multiple-factor predictive models outperform single-factor variants resoundingly for aggregate stocks and for the identified sectoral stocks. This finding is robust, as it is consistent in-sample, out-of-sample, and over different forecast periods.

Furthermore, we examine whether the forecast performance of our multiple-factor model could be improved by extending the model to account for structural breaks. Our results show that the predictability of stock returns for the consumer discretionary, consumer staples, and information technology sectors improves when structural breaks are accounted for in the respective multiple-factor predictive models. However, for aggregate stock returns and the health sector,

accounting for structural breaks in the relevant multiple-factor model tends to generate noise in the model, as it reduces the fitness of the model.

As regards whether accounting for oil price asymmetry improves the forecast performance of our proposed multiple-factor predictive model, our results show that accounting for oil price asymmetry does not contribute significantly to improving the predictability of US stock returns. This is because the multiple-factor predictive model outperforms the asymmetric multiple-factor predictive model for aggregate stock returns, and inconsistent results were observed for relevant sectors such as consumer discretionary, health, and information technology. The multiple-factor model with positive oil price changes outperforms the symmetric multiple-factor predictive model consistently only in the predictive model for stock returns of the consumer staples sector, alone among other sectors accepting the multiple-factor model, such as consumer discretionary, health, and information technology. This implies that accounting for oil price asymmetry in the macroeconomic variable-augmented multiple-factor predictive model will mostly lead to over-fitness of the multiple-factor predictive model for US stock returns.

In summary, our empirical analyses established that the predictability of the oil–US stock predictive model could be improved by capturing the effects of certain domestic macroeconomic factors, such as inflation, interest rate, and economic activity. For aggregate (S&P 500) stock returns, it is not necessary to account for either structural breaks or oil price asymmetry in the multiple-factor model. In the multiple-factor predictive model for stock returns for the consumer discretionary, health, and information technology sectors, accounting for structural breaks is important, but accounting for oil price asymmetry is irrelevant. In the multiple-factor predictive model for stock returns for the consumer staples sector, accounting for both structural break and oil price asymmetry would improve the forecast performance. Conversely, the single-factor predictive model is sufficient for predicting stock returns for the finance and utility sectors, as the multiple-factor predictive model is less efficient in predicting stock returns for these sectors. Our results are robust in-sample and out-of-sample, under 50% and 75% sample sizes, and under short- to medium-term forecast horizons (that is, from one month through six months to 12 months ahead). The fact that our multiple-factor predictive model also outperforms the historical average model further confirms the robustness of our results. However, it may be necessary to pre-test the predictors for persistence, endogeneity, and conditional heteroscedasticity, particularly when modeling with high-frequency series.

References

Ang A. and Bekaert G. (2007). Return predictability: Is it there? Review of Financial Studies 20, 651–707.

Arouri, M. E. and Rault, C. 2010. Causal relationships between oil and stock prices: some new evidence from gulf oil-exporting countries. International Economics, 122, 41-56.

Asteriou, D. and Bashmakova, Y. (2013). Assessing the impact of oil returns on emerging stock markets: A panel data approach for ten Central and Eastern European countries. Energy Economics, 38 204–211.

Baffes, J., Kose, M. A., Ohnsorge, F., and Stocker, M. (2015). The great plunge in oil prices: Causes, consequences and policy responses. World Bank Group Policy Research Note, PRN/15/01.

Bai, J. and Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models, Journal of Applied Econometrics 18(1): 1-22.

Bannigidadmath, D. and Narayan, P. (2015). Stock return predictability and determinants of predictability and profits. Emerging Markets Review 26, 153–173.

Basher, S., Haug, A., and Sadorsky, P. (2012). Oil prices, exchange rates and emerging stock markets. Energy Economics 34, 227–240.

Bondia, R., Ghosh, S. and Kanjilal, K. 2016. International Crude Oil Prices and the Stock Prices of Clean Energy and Technology Companies: Evidence from Non-Linear Cointegration Tests with Unknown Structural Breaks. Energy, 101, 558-565.

Bossaerts P. and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: What do we learn? Review of Financial Studies 12, 405–428.

Bouri, E., Awartani, B., and Maghyereh, A. (2016). Crude oil prices and sectoral stock returns in Jordan around the Arab uprisings of 2010. Energy Economics 56, 205–214.

Brennan M and Xia Y. (2005). Tay's as good as cay. Finance Research Letter 2, 1–14.

Butler A., Grullon G., and Weston J. (2005). Can managers forecast aggregate market return? Journal of Finance 60, 963–986.

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Chang, K. and Yu, S. (2013). Does crude oil price play an important role in explaining stock return behavior? Energy Economics 39, 159–168.

Chen, N.-F., Roll, R., and Ross, S. A. (1986). Economic forces and the stock market. Journal of Business, 383–403.

Chiang, I. E. and Hughen, W. K. (2017). Do oil futures prices predict stock returns? Journal of Banking and Finance 79, 129–141.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Conrad, C., Loch, K., and Rittler, D. (2014). On the macroeconomic determinants of long-term volatilities and correlations in US stock and crude oil markets. Journal of Empirical Finance 29, 26–40.

Cook, S. (2008). Maximum likelihood unit rooting test in the presence of GARCH: A new test with increased power. Communication in Statistics - Simulation and Computation 37(4), 756–765.

Devpura, N., Narayan, P. K., and Sharma, S. S. (2018). Is stock return predictability time-varying? Journal of International Financial Markets, Institutions & Money 52, 152–172.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics 20(1), 134–144.

Enders, W. (2014). Applied Econometric Time Series, 4th Edition, Hoboken, NJ: Wiley.

Faff, R.W. and Brailsford, T. J. (1999). Oil price risk and the Australian stock market. Journal of Energy Finance and Development 4, 69–87.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F. and French, K. R. (1995). Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131–156.

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fama, E. F. and French, K. R. (2017). International tests of a five-factor asset pricing model. Journal of Financial Economics 123, 441–463.

Filis, G., Degiannakis, S., and Floros, C. (2011). Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. International Review of Financial Analysis 20, 152–164.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. Review of Economic Studies 76(2), 669–705.

Goyal A. and Welch I. (2003). Predicting the equity premium with dividend ratios. Management Science 49, 639–654.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. Journal of Political Economy 91(2), 228–248.

Huang, R., Masulis, R., and Stoll, H. (1996). Energy shocks and financial markets. Journal of Futures Market 16, 1–27.

Huang, S., An, H., Gao, X., and Sun, X. (2017). Do oil price asymmetric effects on the stock market persist in multiple time horizons? Applied Energy 185(2), 1799–1808.

Jones, C. and Kaul, G. (1996). Oil and the stock markets. Journal of Finance 51, 463–491.

Jouini, J. (2013). Return and volatility interaction between oil prices and stock markets in Saudi Arabia. Journal of Policy Modelling 35, 1124–1144.

Kang, W., Ratti, R.A. and Yoon, K. 2014. The impact of oil price shocks on U.S. bond market returns. Energy Economics, 44: 248-258.

Kang, W., Ratti, R. A. and Vespignani, J. (2016). The Impact of Oil Price Shocks on the U.S. Stock Market: A Note on the Roles of U.S. And Non-U.S. Oil Production, Economics Letters 145: 176-81.

Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. American Economic Review 99(3), 1053–1069.

Kilian, L. and Park, C. (2009). The impact of oil price shocks on the US stock market. International Economic Review 50, 1267−1287.

Lee, Y. and Chiou, J. (2011). Oil sensitivity and its asymmetric impact on the stock market. Energy 36, 168–174.

Lewellen, J. (2004). Predicting returns with financial ratios. Journal of Financial Economics 74, 209–235.

Li, Q., Cheng, K., and Yang, Z. (2017). Response pattern of stock returns to international oil price shocks: From the perspective of China's oil industrial chain. Applied Energy 185, 1821–1831.

Liu, L., Ma, F., and Wang, Y. (2015). Forecasting excess stock returns with crude oil market data. Energy Economics 48, 316–324.

Makin, A. J., Narayan, P. K., and Narayan, S. (2014). What expenditure does Anglosphere foreign borrowing fund? Journal of International Money and Finance 40, 63–78.

Mankiw, N. G. and Shapiro, M. (1986). Do we reject too often? Small sample properties of tests of rational expectations models. Economic Letters 20, 139–145.

Narayan, P. K. and Bannigidadmath, D. (2015). Are Indian stock returns predictable? Journal of Banking & Finance 58, 506–531.

Narayan, P. K. and Gupta, R. (2015). Has oil price predicted stock returns for over a century? Energy Economics 48, 18–23.

Narayan, P. K. and Liu, R. (2011). Are shocks to commodity prices persistent? Applied Energy 88, 409–416.

Narayan, P. K. and Liu, R. (2013). New evidence on the weak-form efficient market hypothesis. Working Paper. Centre for Financial Econometrics, Deakin University.

Narayan, P. K. and Liu, R. (2015). A unit root model for trending time-series energy variables. Energy Economics 50, 391–402.

Narayan, P. K. and Popp, S. (2010). A new unit root test with two structural breaks in level and slope at unknown time. Journal of Applied Statistics 37(9), 1425–1438.

Narayan, P. K. and Sharma, S. S. (2014). Firm return volatility and economic gains: The role of oil prices. Economic Modelling 38, 142–151.

Narayan, P. K., Liu, R., and Westerlund, J. (2016). A GARCH model for testing market efficiency. Journal of International Financial Markets, Institutions and Money 41, 121–138.

Nelson, C. R. and Kim, M. J. (1993). Predictable Stock Returns: The Role of Small Sample Bias, The Journal of Finance 48(2): 641-61.

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. Journal of Empirical Finance 13(3), 274–315.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics 137(1), 134–161.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. Journal of Econometrics 177(2), 134–152.

Phan, D. H. B., Sharma, S. S., and Narayan, P. K. (2015). Stock return forecasting: Some new evidence. International Review of Financial Analysis 40, 38–51.

276.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate US stock returns. Journal of Financial Econometrics 4(2), 238–274.

Rapach, D. E. and Wohar, M. E. (2008). Forecasting in the Presence of Structural Breaks and Model Uncertainty. Volume 3, Emerald Group Publishing.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Roubaud, D. and Arouri, M. (2018). Oil prices, exchange rates and stock markets under uncertainty and regime-switching. Finance Research Letters, https://doi.org/10.1016/j.frl.2018.02.032.

Sadorsky, P. (1999). Oil price shocks and stock market activity. Energy Economics 21, 449–469.

Salisu A. A, Ademuyiwa, I., and Isah, K. (2018). Revisiting the forecasting accuracy of Phillips curve: The role of oil price. Energy Economics 70(C), 334–356.

Salisu A. A. and Oloko, T. F. (2015). Modelling oil price-US stock nexus: A VARMA-BEKK-AGARCH approach. Energy Economics 50, 1–12.

Salisu, A. A. and Adeleke, A. I. (2016). Further application of Narayan and Liu (2015) unit root model for trending time series. Economic Modelling 55, 305–314.

Salisu, A. A. and Fasanya, I. O. (2013). Modelling oil price volatility with structural breaks. Energy Policy 53, 554–562.

Salisu, A. A. and Isah, K. (2017). Revisiting the oil price and stock market nexus: A nonlinear panel ARDL approach. Economic Modelling 66(C), 258–271.

Salisu, A. A. and Isah, K. (2018). Predicting US inflation: Evidence from a new approach. Economic Modelling 71(C), 134–158.

Salisu, A. A. and Mobolaji, H. (2013). Modeling returns and volatility transmission between oil price and US–Nigeria exchange rate. Energy Economics 39, 169–176.

Salisu, A. A., Ndako, U. B., Oloko, T. F., and Akanni, L. O. (2016). Unit root modeling for trending stock market series. Borsa Istanbul Review 16(2), 82–91.

Salisu, A. A., Swaray, R., and Oloko, T. F. (2017). US stocks in the presence of oil price risk: Large cap vs. small cap. Economics and Business Letters 6(4), 116–124.

Shanken, J. and Weinstein, M. I. (2006). Economic forces and the stock market revisited. Journal of Empirical Finance 13(2),129–44.

Shin, Y., Yu, B. and Greenwood-Nimmo, M. (2014). Modelling Asymmetric Cointegration and Dynamic Multipliers in a Nonlinear Ardl Framework, Festschrift in Honor of Peter Schmidt, pp. 281-314: Springer.

Smyth, R. and Narayan, P. K. (2018). What do we know about oil prices and stock returns? International Review of Financial Analysis 57, 148–156.

Solnik, B. (1987). Using financial prices to test exchange rate models: A note. Journal of Finance 42 (1) 141–149.

Soucek, M. and Todorova, N. (2013). Economic significance of oil price changes on Russian and Chinese stock markets. Applied Financial Economics 23, 561–571.

Stambaugh, R. (1986). Bias in regressions with lagged stochastic regressors. Unpublished Manuscript, University of Chicago, Chicago, IL.

Stambaugh, R. F. (1999). Predictive regressions. Journal of Financial Economics 54, 375–421.

Swaray, R. and Salisu, A. A. (2018b). A Firm-Level Analysis of the Upstream-Downstream Dichotomy in the Oil-Stock Nexus, Global Finance Journal 37: 199-218.

Tsagkanos, A. and Siriopoulos, C. (2015). Stock markets and industrial production in north and south of Euro-zone: Asymmetric effects via threshold cointegration approach. Journal of Economic Asymmetries 12, 162–172.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21(4): 145–1508.

Westerlund, J. and Narayan, P. K. (2012). Does the choice of estimator matter when forecasting returns? Journal of Banking and Finance 36, 2632–2640.

Westerlund, J. and Narayan, P. K. (2015). Testing for predictability in conditionally hetoroscedastic stock returns. Journal of Financial Econometrics 13, 342–375.

Wu, S.-J. and Lee, W.-M. (2015). Predicting severe simultaneous bear stock markets using macroeconomic variables as leading indicators. Finance Research Letters 13, 196–204.

Yu, H., Du, D., Fang, L., and Yan, P. (2018). Risk contribution of crude oil to industry stock returns. International Review of Economics & Finance, https://doi.org/10.1016/j.iref.2018.03.009.

Appendix

Dealing with endogeneity, persistence, and conditional heteroscedasticity in the

predictive model for stock returns

To test for endogeneity and persistence and to determine an estimable predictive regression model that accounts for these features, we follow the procedure as detailed in Lewellen (2004) and Westerlund and Narayan (2012, 2015). Let us assume predictive model 1A for the stock price, as previously specified:

$$r_t = \alpha + \beta x_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right) \tag{1A}$$

where r_t is as previously defined, and x_t is a potential predictor variable of stock price restricted to oil price, output, and inflation. We can also assume an AR(1) process for x_t (as in 2A) (see also Stambaugh, 1986, 1999; Mankiw and Shapiro, 1986; Nelson and Kim, 1993; Lewellen, 2004; Narayan and Gupta, 2014):

$$x_t = \phi + \rho x_{t-1} + \varepsilon_{x,t}; \quad \varepsilon_{x,t} \sim N\left(0, \sigma_{\varepsilon_{x,t}}^2\right) \tag{2A}$$

Both ε_t and $\varepsilon_{x,t}$ are expected to be correlated for endogeneity bias to pose any serious concern on the outcome of the forecast and $\hat{\rho} < 1$, an assumption that is required for stationarity. The null hypothesis of no predictability is given as $H_0 : \hat{\beta} = 0$. However, in the presence of persistence and endogeneity, Lewellen (2004) shows that $\hat{\beta}$ exhibits a bias that can be expressed as $E\left(\hat{\beta} - \beta\right) = \gamma\left(\hat{\rho} - \rho\right)$. The direction of bias (i.e., upward or downward) is determined by γ and $\hat{\rho}$. If, for instance, γ is positive (based on the nature of relationship between ε_t and $\varepsilon_{x,t}$), a downward bias in $\hat{\rho}$ causes a downward bias in $\hat{\beta}$. Since x_t is expected to be stationary for any meaningful forecast to be conducted, the bias in $\hat{\beta}$ is at most $\gamma\left(\hat{\rho} - 1\right)$. In the absence of persistence, the bias disappears (although there is still potential correlation between the two errors because of endogeneity bias). However, if there is persistence, $\hat{\rho} \neq 0$ and any predictive model that ignores the information in $\hat{\rho}$ tends to understate the predictive power of the predictor(s). As

demonstrated in the literature, any predictive regression model that accounts for the information in $\hat{\rho}$ significantly strengthens its forecasting accuracy (for a review, see Lewellen, 2004).

To test for the degree of persistence with the null hypothesis of $H_0 : \hat{\rho}=1$, we estimate 2A with the OLS estimator, and the first lag autoregressive coefficient is used to determine the extent of persistence. The closer this coefficient is to 1, the higher the degree of persistence. To test and capture the endogeneity effects, we follow the approach of Westerlund and Narayan (2012, 2015), which establishes the relationship between the two errors $\left(\varepsilon_t \text{ and } \varepsilon_{x,t}\right)$ as in 3A:

$$\varepsilon_t = \gamma \varepsilon_{x,t} + \eta_t \tag{3A}$$

where ε_t and $\varepsilon_{x,t}$ are defined as before, η_t has a zero mean and variance σ_η^2, and $E\left(\varepsilon_{x,t}\eta_t\right)=0$. In other words, η_t is the remainder error of ε_t after controlling for $\varepsilon_{x,t}$, and therefore, by construct, $\varepsilon_{x,t}$ and η_t are not expected to be correlated. The null hypothesis of no endogeneity bias is given as $H_0 : \gamma=0$, which can be tested by replacing the error terms in 3A by their corresponding residuals obtained from the estimation of 1A and 2A. By way of substitution and re-arrangement, we can rewrite equation 11 as in 4A:

$$r_t = \delta + \beta x_{t-1} + \gamma\left(x_t - \rho x_{t-1}\right) + \eta_t \tag{4A}$$

where $\delta = \alpha - \gamma\phi$. In order to correct for the inherent bias in β, Lewellen (2004) suggests bias-adjusted OLS estimator of β, which is described in 5A.

$$\hat{\beta}_{adj} = \hat{\beta} - \gamma\left(\hat{\rho} - \rho\right) \tag{5A}$$

Since ρ is unknown, Lewellen (2004) suggests we can put a lower bound on $\left(\hat{\rho} - \rho\right)$ by assuming that $\rho \approx 1$ (in line with the null hypothesis for testing for unit root), while $\hat{\rho}$ is determined from 2A. We follow the same approach in this paper to estimate $\hat{\beta}_{adj}$. Note that, in the absence of persistence and endogeneity effects, $\hat{\beta}_{adj} = \hat{\beta}$, since $\gamma\left(\hat{\rho} - \rho\right)=0$ and, by implication, $\gamma\left(x_t - \rho x_{t-1}\right)=0$. Therefore, in the absence of persistence and endogeneity effects, 4A reduces to

2A, and the standard OLS is valid for estimation in that instance. The case is similar if there is no evidence of endogeneity but the predictor variable still exhibits persistence; provided the order of integration for the predictor does not exceed $I(1)$, 1A is a valid predictive regression model. However, if there is no evidence of persistence but endogeneity bias is evident, 4A becomes as in 6A:

$$r_t = \delta + \beta x_{t-1} + \gamma x_t + \eta_t \tag{6A}$$

where $\hat{\rho} \approx 0$ and, as a consequence, bias-adjusted OLS estimator of β in 5A adjusts to $\hat{\beta}_{adj} = \hat{\beta} + \gamma \rho$, and where $\rho \approx 1$, $\hat{\beta}_{adj} = \hat{\beta} + \gamma$. In this case, the adjustment in $\hat{\beta}_{adj}$ is determined by the sign on γ. For instance, a negative correlation between the two errors captured by γ implies that $\hat{\beta}$ adjusts upwards when $\hat{\rho} \approx 0$.

In addition to endogeneity and persistence effects, dealing with conditional heteroscedasticity becomes necessary for financial markets, such as stock markets, that are susceptible to shocks and volatility. We follow the approach proposed by Westerlund and Narayan (2012, 2015) to account for conditional heteroscedasticity in the model. Westerlund and Narayan (2012) propose a FQGLS estimator, which exploits the information contained in the conditional heteroscedastic variance of the regression residuals in order to generate more precise estimates. Their estimator assumes the regression error, that is η_t, follows an autoregressive conditional heteroscedastic (ARCH) structure $\hat{\sigma}_{\eta,t}^2 = \mu + \sum_{i=1}^{q} \varphi_i \hat{\eta}_{t-i}^2$, and the resulting $\hat{\sigma}_{\eta,t}^2$ can be used as a weight in the predictive model. The estimation of the weighted predictive model by OLS is described as the FQGLS estimator. Essentially, the GLS-based t-statistic for testing $\beta = 0$ is given in 7A:

$$t_{FQGLS} = \frac{\sum_{t=q_m+2}^{T} \tau_t^2 x_{t-1}^d r_t^d}{\sqrt{\sum_{t=q_m+2}^{T} \tau_t^2 \left(x_{t-1}^d\right)^2}} \tag{7A}$$

where $\tau_t = 1/\sigma_{\eta,t}$ is used in weighting all the data in the predictive model, and $x_t^d = x_t - \sum_{s=2}^{T} x_t/T$. In the case of two predictors, for example, we express the model as in 8A-A:

$$r_t = \alpha + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \varepsilon_t \tag{8A-A}$$

where x_1 and x_2 are two potential predictors of stock price. The error term ε_t has a mean of zero and variance of σ_ε^2. As in the previous analyses, we can assume a first-order autoregressive model for x_1 and x_2 as in 8A-B and 8A-C:

$$x_{1t} = \phi_1 + \rho_1 x_{1,t-1} + \varepsilon_{x_1,t} \qquad \text{(8A-B)}$$

$$x_{2t} = \phi_2 + \rho_2 x_{2,t-1} + \varepsilon_{x_2,t} \qquad \text{(8A-C)}$$

where $\varepsilon_{x_1,t}$ and $\varepsilon_{x_2,t}$ are assumed to have zero mean values and $\sigma_{\varepsilon_{x1}}^2$ and $\sigma_{\varepsilon_{x2}}^2$ as their respective variances. The null hypothesis of perfect persistence, $H_0 : \rho = 1$, is tested against the alternative of $H_1 : \rho < 1$. To capture endogeneity effects, we assume correlation 9A between ε_t and $\varepsilon_{x_1,t}$ and $\varepsilon_{x_2,t}$:

$$\varepsilon_t = \gamma_1 \varepsilon_{x_1,t} + \gamma_2 \varepsilon_{x_2,t} + \eta_t \qquad \text{(9A)}$$

Since η_t is the remainder error of $\varepsilon_{r,t}$ after accounting for $\varepsilon_{x_1,t}$ and $\varepsilon_{x_2,t}$, it is important to assume that $E\left(\varepsilon_{x_1,t}\eta_t\right) = E\left(\varepsilon_{x_2,t}\eta_t\right) = 0$. To implement 9A, it is also important to assume the absence of correlation between $\varepsilon_{x_1,t}$ and $\varepsilon_{x_2,t}$ which can easily be verified by expressing one against the other (say, $\varepsilon_{x_1,t} = \delta\varepsilon_{x_2,t} + v_t$) and testing for the significance of the relevant parameter δ. The null hypothesis of no endogeneity in 9A, $H_0 : \gamma_1, \gamma_2 = 0$, is tested against the alternative of $H_1 : \gamma_1, \gamma_2 \neq 0$.
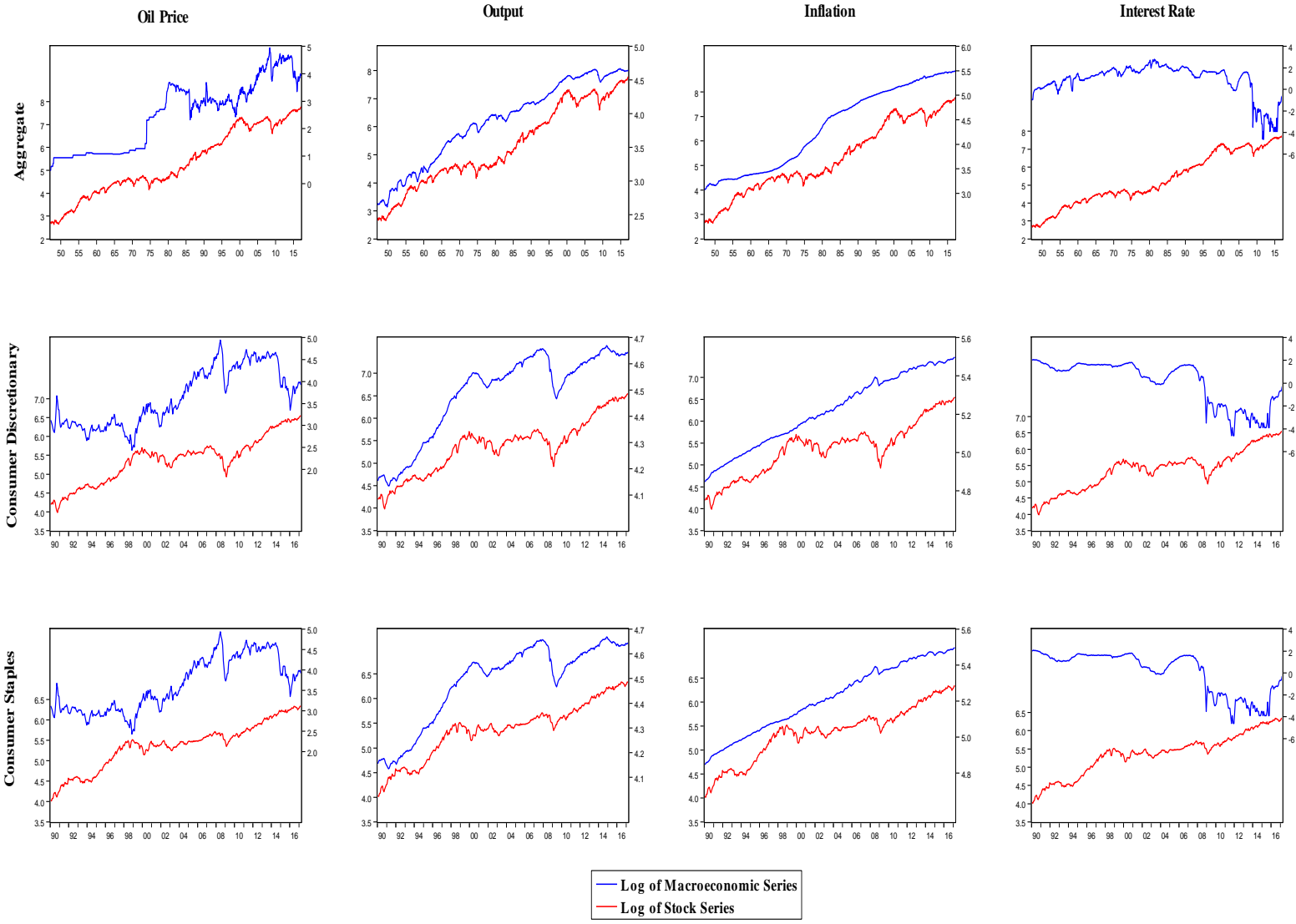
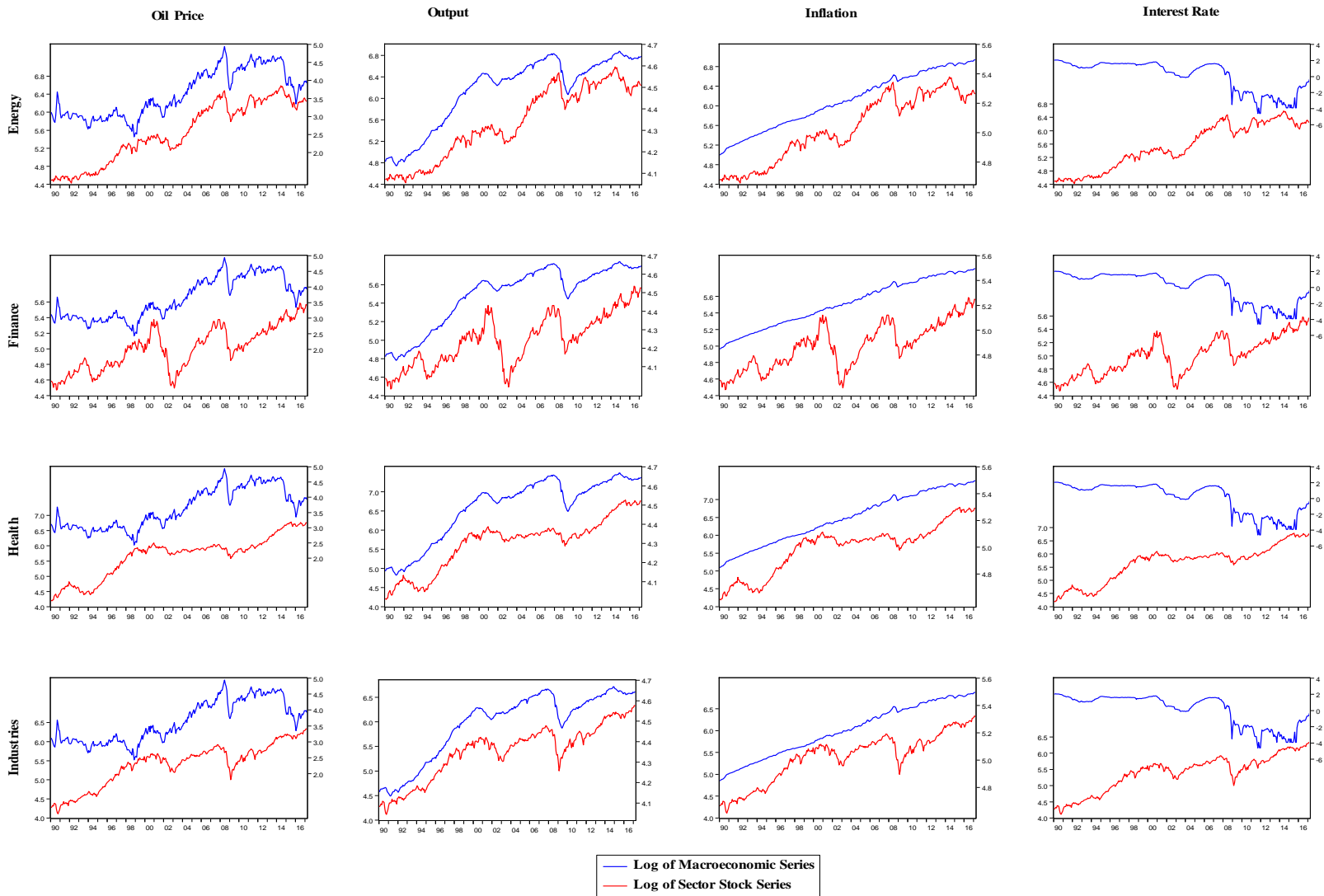By substitution and further simplification of equation 9A, we can rewrite equation 7A as in 10A:

$$\pi_t = \omega + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \gamma_1 \left(x_{1t} - \rho_1 x_{1,t-1} \right) + \gamma_2 \left(x_{2t} - \rho_2 x_{2,t-1} \right) + \eta_t \qquad \text{(10A)}$$

where $\omega = \alpha - \phi_1\gamma_1 - \phi_2\gamma_2$, and the corresponding equation for the bias-corrected OLS estimates for the βs is given in 11A:

$$\pi_t = \omega + \beta_{1,adj} x_{1,t-1} + \beta_{2,adj} x_{2,t-1} + \gamma_1 \left(x_{1t} - \rho_1 x_{1,t-1} \right) + \gamma_2 \left(x_{2t} - \rho_2 x_{2,t-1} \right) + \eta_t \qquad \text{(11A)}$$

where $\hat{\beta}_{1,adj} = \hat{\beta}_1 - \hat{\gamma}_1\left(\hat{\rho}_1 - 1\right)$, and $\hat{\beta}_{2,adj} = \hat{\beta}_2 - \hat{\gamma}_2\left(\hat{\rho}_2 - 1\right)$. It is easy to verify that 11A reduces to 1A in the absence of persistence and endogeneity effects.

Oil Price **Output** **Inflation** **Interest Rate**

Aggregate

Consumer Discretionary

Consumer Staples

Log of Macroeconomic Series
Log of Stock Series

47

	Oil Price	Output	Inflation	Interest Rate
Energy				
Finance				
Health				
Industries				

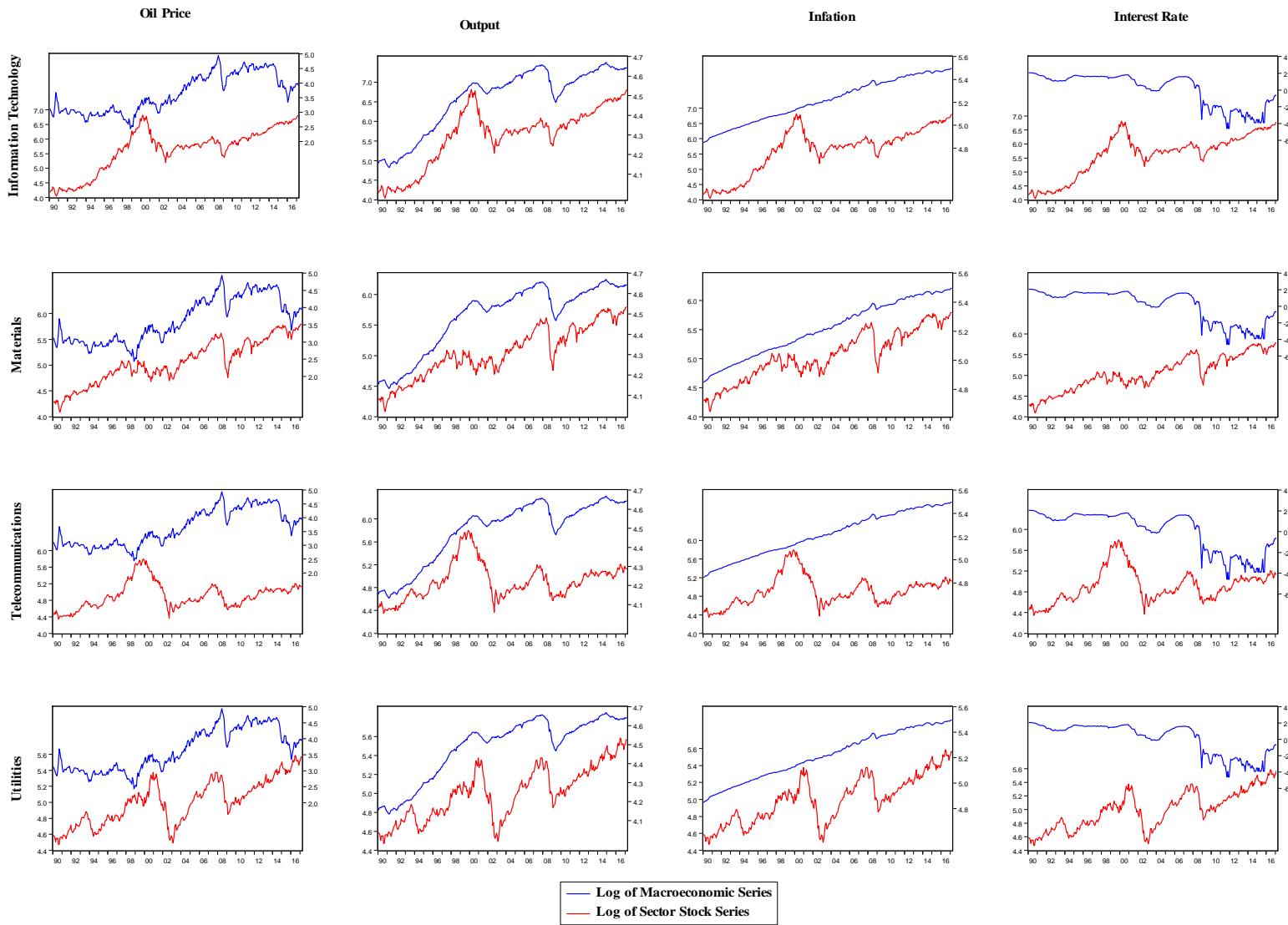Log of Macroeconomic Series
Log of Sector Stock Series

48

Figure A1. Graphical representation of trends in stock market performance and macroeconomic variables.

Table A1

Structural breaks in the multiple-factor model using the Bai–Perron (2003) test.

Stock returns	50% sample		75% sample	
	D1	D2	Z1	Z2
Aggregate	264	-	264	437
Cons_disc	140	-	210	-
Cons_staples	-	-	125	-
Energy	-	-	182	-
Finance	138	-	133	-
Health	-	-	126	-
Industries	100	-	210	-
Info_tech	129	-	-	-
Materials	-	-	210	-
Telecom	100	-	115	154
Utility	138	-	133	-