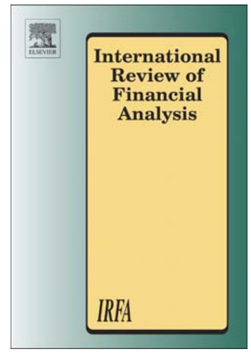
Accepted Manuscript



Stock Return Forecasting: Some New Evidence

Dinh Hoang Bach Phan, Susan Sunila Sharma, Paresh Kumar Narayan

PII:	S1057-5219(15)00084-8
DOI:	doi: 10.1016/j.irfa.2015.05.002
Reference:	FINANA 841
To appear in:	*International Review of Financial Analysis*
Received date:	28 November 2014
Revised date:	10 March 2015
Accepted date:	13 May 2015

Please cite this article as: Phan, D.H.B., Sharma, S.S. & Narayan, P.K., Stock Return Forecasting: Some New Evidence, *International Review of Financial Analysis* (2015), doi: 10.1016/j.irfa.2015.05.002

Stock Return Forecasting: Some New Evidence

Dinh Hoang Bach Phan, Susan Sunila Sharma, Paresh Kumar Narayan

Mailing Address

Dinh Hoang Bach Phan
Centre for Economics and Financial Econometrics Research
Deakin University
221 Burwood Highway
Burwood, Victoria 3125
Australia
Email: dphan@deakin.edu.au

Stock Return Forecasting: Some New Evidence

ABSTRACT

This paper makes three contributions to the literature on forecasting stock returns. First, unlike the extant literature on oil price and stock returns, we focus on out-of-sample forecasting of returns. We show that the ability of the oil price to forecast stock returns depends not only on the data frequency used but also on the estimator. Second, out-of-sample forecasting of returns is sector-dependent, suggesting that oil price is relatively more important for some sectors than others. Third, we examine the determinants of out-of-sample predictability for each sector using industry characteristics and find strong evidence that return predictability has links to certain industry characteristics, such as book-to-market ratio, dividend yield, size, price earnings ratio, and trading volume.

Keywords: *Stock returns; Oil price; Predictability; Forecasting; Out-of-sample.*

1. Introduction

Several studies show that financial ratios and macro variables, such as dividend-price, price-earnings, dividend pay-out, and book to market ratios, inflation rate, interest rates, aggregate output predict stock returns (see, *inter alia*, Fama, 1981; Campbell, 1987; Fama and French, 1988,1989; Campbell and Shiller, 1988a,b; Kothari and Shanken, 1997; Pontiff and Schall, 1998; Lamont, 1998; Rapach *et al.*, 2005). However, the literature finds relatively limited evidence of predictability using out-of-sample tests. The general conclusion is that the evidence for stock return (US market) predictability is predominantly in-sample; and it is not robust to out-of-sample evaluations (see, for example, Bossaerts and Hillion, 1999; Goyal and Welch, 2003; Brennan and Xia, 2005; Butler *et al*., 2005; and Ang and Bekaert, 2007). To confirm the lack of out-of-sample evidence, Welch and Goyal (2008) evaluate the performance of several predictive regression models for the equity premium using a wide range of financial ratios and macroeconomic predictors. They find that the out-of-sample stock return forecast fails to beat the simple historical average benchmark forecast. The historical average forecast assumes that the coefficients of the predictors are equal to zero, implying that the information from economic predictors is not useful for predicting stock returns. They also test the predictive power of the predictors by including all predictors in a single model, which they refer to as the “kitchen sink” model. However, the performance of this model is even worse due to in-sample over-fitting. This is not surprising, as it is well-known that highly parameterised models typically perform very poorly in out-of-sample evaluations (Welch and Goyal, 2008). They conclude that the predictive regression models are not stable and are unable to beat the historical average.

The findings from Welch and Goyal (2008) have instigated further responses. Campbell and Thompson (2008), for instance, introduce an economically-motivated restriction model approach, which restricts the sign of the predictor coefficients and forecasted returns. According to their approach, the predictor coefficients must have the theoretically expected sign and the forecasted returns must be positive or equal to zero. By applying these restrictions sequentially and jointly, they show that the restricted model substantially improves the forecast ability of the predictive regression model in out-of-sample tests to the extent that it is able to beat the historical average forecasting model. The restricted models also generate economically meaningful utility gains for mean-variance investors. Rapach *et al.* (2010) propose improving predictability by combining numerous variables in a predictive regression. The main benefit of this approach is that it incorporates information from numerous economic variables while reducing forecast volatility. They show that the combination predictive regression, which includes 15 economic variables, can beat the historical average model in out-of-sample forecasting of the stock returns in different sample periods.

On the other hand, several studies extend the literature by proposing econometric models that deal with issues of bias and inefficiency in stock return predictive regression models. Westerlund and Narayan (2015a), for instance, point out that the forecasting regression may face a number of potential issues, including heteroskedasticity, predictor endogeneity, and persistency. Many predictors are persistent and could lead to biased coefficients in predictive regressions if the innovation of the predictor is correlated with return innovations (Nelson and Kim, 1993; Stambaugh, 1999). In addition, one of the well-documented features of financial time series data is that the returns are highly heteroskedastic, which is another source of bias and inefficiency in the predictive regression models.

We use daily, weekly, and monthly S&P500 indices over the period 4 January 1988 to 31 December 2012. Three forecasting approaches for the out-of-sample stock returns based on the crude oil price are utilised. They are the Ordinary Least Squares (OLS), adjusted OLS (AOLS, Lewellen, 2004) and the Feasible Generalised Least Squares (FGLS, Westerlund and Narayan, 2015a). In-sample periods of 25%, 50% and 75% are chosen. This is important since the literature has shown that the results can vary depending on the in-sample periods, particularly in the finite sample sizes (Rozeff, 1984; Fama and French, 1988; Lettau and Nieuwerburgh, 2008; Boudoukh *et al.,* 2008; Narayan *et al.,* 2013). The evaluation statistics for forecasting performance are out-of-sample R-squared (R^2_{OS}) and Diebold-Mariano test for non-nested models and the Mean Square Forecast Error (MSFE) adjusted, which is introduced by Clark and West (2007), for nested-models.

Briefly foreshadowing the main results, our findings are as follows. First, the FGLS approach is superior to the OLS and AOLS approaches in forecasting stock returns, giving credence to our idea of accounting for, in particular, the heteroskedasticity feature of high frequency time-series financial data. The key implication of this finding is that the choice of the estimator in forecasting returns is important and deserves particular attention. While this message is consistent with Westerlund and Narayan (2012), we add to this literature by showing that the relevance of the estimator matters not only when using traditional predictors of stock returns but also when using non-traditional stock return predictors, such as oil prices. Our study, therefore, adds to the robustness of the role estimators tend to play in forecasting.

Second, a crude oil price-based forecasting model estimated by the FGLS approach (but not OLS and AOLS approaches) generally beats the historical average model in the out-of-sample forecasting evaluation. Moreover, a positive utility gain for a mean-variance investor is observed, suggesting that the oil price-based forecasting model estimated by the FGLS approach is not only statistically significant but also economically superior. The key implication of this finding is that while traditional financial ratios and selected macroeconomic variables have been found to predict stock returns, non-traditional variables which have serious economic repercussions, such as oil prices as shown in Narayan and Sharma (2011), should not be ignored[1]. In our study, oil price is not only statistically related to stock returns but there is a strong economic connection as well. This finding also points toward a need to explore other non-traditional predictors of returns, such as gold prices, amongst others.

We also document that the outperformance of the FGLS model is data frequency-dependent. The key implication of this finding is that data frequency matters regarding forecasting outcomes and therefore should not be ignored. This message supports those highlighted by Narayan and Sharma (2015) regarding the importance of data frequency in hypotheses tests, and Narayan *et al.* (2013) and Narayan *et al.* (2015) regarding the importance of data frequency when it comes to the profitability of commodities. We add to these studies by showing that data frequency matters beyond commodity profits and hypotheses tests to stock return predictability.

[1] Amongst recent work that utilises non-traditional predictors of stock returns are: Narayan *et al.* (2014b), who use mutual funds; Narayan *et al.* (2014a), who use institutional quality; and Narayan *et al.* (2015b), who use quality of governance. The main message emerging from these studies is that stock return predictability exists beyond the traditional predictors of stock returns.

Finally, the results suggest that sector characteristics such as book-to-market ratio, dividend yield, size, price-earnings ratio, and trading volume significantly determine sectoral return predictability. There is an important message emerging from these results. Traditionally, the stock return predictability literature has shown that a range of variables—both financial ratios and macroeconomic indicators—are able to predict stock returns to some extent at least; for a recent analysis and related references, see Westerlund and Narayan (2015b). Stock return predictability has been shown to be stock characteristic (size and trading volume) dependent[2]. The key point here is that these variables are able to predict stock returns so they can potentially also determine stock return predictability over time. This is because these variables have information content which is useful for the stock market. Whether or not these variables determine stock return predictability, therefore, is an empirical issue and we ascertain this empirically.

The remainder of the paper is organised as follows. The motivation and contribution of this paper are presented in section 2 while the methodology and statistics are discussed in section 3. Section 4 discusses our main findings and concluding remarks are provided in section 5.

2. Motivation and contribution

In the stock return predictability tests, the OLS is the most commonly-used estimator. Its main limitation is that it ignores the fundamental data issues, such as persistency, endogeneity, and heteroskedasticity. Lewellen (2004) introduces the bias-adjusted OLS that accounts for the predictor endogeneity and persistency issues. As a result, the bias does not appear in the forecasts using the AOLS estimator, however, the inefficiency still exists. In response to this, Westerlund and Narayan (2012, 2015a) propose a FGLS estimator that accounts for not only the endogeneity and persistency of the predictor variable but also the heteroskedasticity in the predictive regression model. In fact, the results from preliminary analysis show that the aforementioned issues exist in our data sample. Hence, the FGLS model is expected to provide the best performance in forecasting the stock returns. Despite the large volume of studies in stock returns forecasting, there are very few papers which employ either the AOLS or the FGLS approach for out-of-sample forecasting. With regard to the FGLS estimator, only Westerlund and Narayan (2012) have applied this method first-hand with financial ratios to forecast US stock returns, showing that the FGLS approach outperforms the competitor tests.

This paper employs the crude oil price as a predictor of stock returns. There are two reasons why we choose oil price as a predictor. First, there is a rich volume of studies that use oil price to study its effects on stock returns. What appeals to us most is that the literature reports mixed evidence on how oil price affects stock returns. There are some studies that report a negative relationship between crude oil price changes and aggregate stock market returns (Jones and Kaul, 1996; Sadorsky, 1999; Park and Ratti, 2008; Driesprong *et al*., 2008; and Miller and Ratti, 2009); some studies (see, for instance, Chen *et al*., 1986; Huang *et al*., 1996; and Wei, 2003) document no statistically significant effect of the crude oil price on stock returns; while others show that oil price

[2] Several studies explain the expected relationship between financial ratios and stocks returns. Briefly, book-to-market ratio, dividend yield and trading volume are expected to have a positive effect on returns whereas market capitalisation and price-earnings ratio are expected to have a negative effect on returns. For specific explanations, see Ball (1978), Llorente *et al.* (2002), Banz (1981), Fama and French (1988, 1991), and Lewellen (2004).

has a positive effect on stock prices (see Narayan and Narayan, 2010; Narayan and Sharma, 2011; Gjerde and Saettem, 1999; Park and Ratti, 2008). The main feature of these studies is that they only focus on in-sample predictability while nothing is known about how well the oil price predicts out-of-sample stock returns. Tashman (2000) argues that forecasting methods should be assessed for accuracy using out-of-sample tests rather than goodness-of-fit to past data (in-sample tests)[3]. In addition, Welch and Goyal (2008) state that a predictor must provide a good out-of-sample predictive performance in order to be used by an investor. The out-of-sample forecasting analysis would be more relevant for investors as they are required to take decisions in real time.

Our second motivation for choosing oil price has roots in the broader literature on the effects of oil prices on economic growth. Following Hamilton (1983), it is now well accepted that oil price has a negative effect on economic growth. A large literature also shows that higher economic growth stimulates stock market activities. It follows, therefore, that a rise in oil price, which negatively affects economic growth, will in turn have a negative effect on the stock market (Narayan and Sharma, 2011). While this relationship was generally accepted for the United States in particular, a recent study by Narayan *et al.* (2014c) shows the relevance of oil prices for as many of 45 developed and developing countries. There are many reasons for why oil price affects the macroeconomy and therefore economic growth. The first reason is that oil is used as an input in the production process. When oil prices rise generally the cost of production increases. That Hamilton (1983) points out that seven of the eight post-war recessions in the United States has preceded sharp rises in oil prices is not surprising. This implies that oil is almost always at the purview of policymakers. The second reason is related to the treatment of crude oil as an asset class. Several studies now show how investing in oil can be profitable; see, *inter alia*, Narayan et al. (2013) and Narayan et al. (2015). Oil investing therefore is at the heart of commodity-based investment. This implies that oil is in the purview of investors. On the whole the focus and emphasis on oil prices is unrelenting.

[3] One of the reasons for an out-of-sample test preference is that in-sample errors are likely to understate forecasting errors for a given forecasting method. The selected in-sample test approach is designed to forecast the historical data but the nuances of past history may not remain in the future and the nuances of the future may not have revealed themselves in the historical data (Tashman, 2000). Furthermore, methods selected by best in-sample fit may not be the best for out-of-sample forecasting. A large number of predictors, which are successful in in-sample predicting, cannot pass out-of-sample forecasting in stock return predictability literature (e.g., Bossaerts and Hillion, 1999; Goyal and Welch, 2003; Welch and Goyal, 2008).

Generally speaking, there is simply no theory that dictates the choice of data frequency (see Narayan and Sharma, 2015). In the stock returns forecasting literature, different data frequencies have been used. This is understandable because forecasting is undertaken for a purpose which dictates the choice of data frequency. For example, policy makers may be interested in knowing quarterly forecasts of financial data in order to determine expected inflation and short rates. For this purpose, quarterly data may be used. Stocks of service industries, such as tourism, for instance, may be interested in monthly forecasts of stock returns since visitor arrivals figures in most countries are released monthly. In this case, the preference in forecasting will be for monthly data. When it comes to stock return predictability, monthly, weekly and annual data have been used. Therefore, it seems that one should confirm the robustness of the forecasting performance by using data at different frequencies. This is not a trivial matter because earlier studies show that the choice of data frequency does matter. For a discussion on this, see Narayan and Sharma (2015).

Most studies based on out-of-sample forecasting of stock returns use low-frequency data such as annual data (Welch and Goyal, 2003), quarterly data (Rapach *et al.,* 2010), or monthly data (Bossaerts and Hillion, 1999; Welch and Goyal, 2008; Campbell and Thompson, 2008; Rapach *et al.,* 2011; Kong *et al.,* 2011; Westerlund and Narayan, 2012). This paper investigates stock return predictability using a crude oil predictor that has higher frequency data available, which captures better insights due to the ability to forecast at short-horizon for stock return predictability.

Unlike the literature on the market level, the studies on out-of-sample predictability at market component level are very limited (see Rapach *et al*., 2011; Kong *et al.,* 2011)[4]. The limitation of the predictability literature at aggregate market level is that it assumes that the predictability of the component in the market is homogenous. This is a strong assumption and needs to be tested by specifically examining the predictability of individual sectors[5]. Rapach *et al*. (2011) and Kong *et al.* (2011) show that the predictability of market components is heterogeneous. That is some sectors are predictable others are not. Thus, the determinants of predictability among sectors are debatable and worthy of investigation. While studies like Avramov and Chordia (2006) examine the determinants of predictability they do so from the point of view of the relative importance of time-varying alpha and market beta explaining predictability. This is based on an in-sample test. To-date, no attempt has been made to explore the time-series determinants of predictability using an out-of-

[4]Rapach *et al*. (2011) compute the diffusion index forecasts based on 14 economic variables and 33 lagged industry returns to forecast the return of 33 industry portfolios from French's database. They show substantial differences in the degree of return predictability in terms of sign, magnitude, and significance of the R^2_{OS} across portfolios. Kong *et al.* (2011) examine 25 Fama-French size and value sorted portfolios using economic variables and portfolio lagged return. The variables significantly predict industry stock returns in most of the cases. However, the degree of predictability power is different among industry portfolios. The predictability is stronger for the portfolio with low market capitalization and high book-to-market ratio.

[5]According to Rapach *et al*. (2011), investigating the predictability at market component has an important application in a dynamic asset pricing model (Stambaugh, 1983; Campbell, 1987; Connor and Korajczyk, 1989; Ferson and Harvey, 1991; and Kirby, 1998). From the point of view of portfolio management, identifying the predictability of individual sectors is important for capital allocation among sectors and sector-rotation strategies. Since the predictability of sectors is divergent, a portfolio manager may want to allocate more or shift investment assets from one sector to another that has better forecasting performance. The investment strategies that ignore the industry portfolio return predictability are substantially outperformed by the strategies that consider the predictability, which provides higher Sharpe ratios, generating sizeable alphas and exhibiting timing ability (Kong *et al.,* 2011).

sample test. We do this. This paper fills the gap in the literature by examining the determinants of out-of-sample predictability using a number of industry characteristics, such as book-to-market ratio, dividend yield, market capitalisation, price-earnings ratio, and trading volume. These variables are selected due to their impact on stock returns. The predictive ability of the book-to-market ratio, dividend yield, and price-earnings ratio on stock return has been extensively proved in previous studies (see, *inter alia*, Campbell and Thompson, 2008; Rapach *et al.,* 2010; Westerlund and Narayan, 2012). While examining the return predictability of industry return sorted by size and book-to-market ratio, Kong *et al.* (2011) find that the predictability is stronger for the portfolio with low market capitalisation and a high book-to-market ratio. In addition, a positive relationship between trading volume and stock return is consistently found in previous studies (e.g., Clark, 1973; Copeland, 1976; Tauchen and Pitts, 1983; Karpoff, 1987; Jones *et al.*, 1994; Foster, 1995; Wang and Yau, 2000; Chen *et al.,* 2001). Due to the influence of those variables on stock return, we also believe that they are able to explain stock return predictability.

Our contributions to the literature are summarised as follows. First, we contribute further insights for the crude oil and stock market relationship literature by employing an out-of-sample analysis utilising a recently developed FGLS estimator to show how well the oil price forecasts stock returns. Second, we test the robustness of forecasting performance by using different forecasting estimators at different data frequencies and find that the results depend not only on the data frequency used but also on the estimator. Third, we confirm that the out-of-sample stock return predictability is sector-dependent and links to certain industry characteristics, such as book-to-market ratio, dividend yield, size, price-earnings ratio, and trading volume.

3. Methodology

3.1. Predictive regressions

Following the previous studies (Campbell and Thompson, 2008; Rapach *et al*., 2010; Welch and Goyal, 2008; and Westerlund and Narayan, 2012), the stock return predictability is tested based on the following predictive regression model:

$$r_{t+h} = \alpha + \beta x_t + \epsilon_{t+h} \tag{1}$$

where r_{t+h} is the excess stock return, measured as stock returns in excess of the risk-free rate which we proxy using the three-month Treasury bill rate, at time t+h, and x_t is the crude oil price which is believed to be able to predict h-period-ahead stock returns.[6] The predictive ability of x_t can be tested under the null hypothesis that $\beta = 0$ (no predictability) and the alternative is $\beta \neq 0$, in which case x_t does predict future stock returns. As much as this model is simple to estimate, it also has a fair share of drawbacks. This model, for instance, suffers from the issues of persistency,

[6] An issue with the specification in Equation (1), as also pointed out by one referee of this journal, is that it ignores many other potential predictors of returns. The bivariate predictive regression model has been preferred because when multiple predictors are included there is the statistical issue of multicollinearity. Therefore, as can be found from the large volume of stock return predictability literature cited here, researchers tend to engage in a horserace amongst all potential predictors, one-by-one. Indeed the econometric developments toward testing the null hypothesis of no predictability has taken place within this bivariate empirical framework. We, therefore, keep to this tradition of using bivariate predictive regression models. Herein, perhaps, lies avenue(s) for future econometric theory work with respect to using multiple predictors in forecasting stocks returns.

endogeneity, and heteroskedasticity. The bias and inefficiency of the predictor coefficient due to the aforementioned issues are discussed in the studies of Stambaugh (1999), Lewellen (2004), and Westerlund and Narayan (2015a).

Assume the predictor in Equation (1) follows an AR (1) process, as below:

$$x_{t+1} = \mu(1-\rho) + \rho x_t + \varepsilon_{t+1} \tag{2}$$

$$\epsilon_t = \gamma\varepsilon_t + \eta_t \tag{3}$$

where $|\rho| \leq 1$. The ϵ_t and ε_t are independently and identically distributed disturbance terms and are assumed to have mean zero and to be uncorrelated to each other (which means $\gamma = 0$). However, this assumption does not always hold, rendering the predictor endogenous, which means the OLS estimator of β is no longer unbiased. In addition, predictors can also be highly persistent. This dual feature of the predictor can induce a small-sample bias in the OLS estimator (see, Stambaugh, 1999). Stambaugh (1999) finds that the coefficient of the predictor is biased by $-\gamma(1+3\rho)/T$ when the predictor variable is not exogenous, such that $\gamma \neq 0$. This bias, however, decreases with the increase in sample size.

Lewellen (2004) argues that although the bias correction of Stambaugh (1999) is generally appropriate, it can substantially understate the forecasting power, particularly when the predictor is highly persistent. Lewellen (2004) takes into account the information from the predictor autoregressive regression regarding $(\hat{\rho} - \rho)$, which Stambaugh (1999) does not. Stambaugh's approach can be appropriate when the value of ρ is small because it contains little information. In response, Lewellen (2004) proposes a bias-adjusted estimator that can obviate the bias inherent in the OLS predictive regression model. The idea is to make Equation (1) conditional on ϵ_t by substituting Equations (2) and (3):

$$r_{t+h} = \theta + \beta x_t + \gamma(x_{t+h} - \rho x_{t+h-1}) + \eta_{t+h} \tag{4}$$

where $\theta = \alpha - \mu(1-\rho)$. As a result, we can remove the correlation between the regression errors. However, the regression cannot be feasible as ρ is unknown, therefore, Lewellen (2004) replaces the true value ρ by a guess, named ρ_0. Therefore, Equation (4) can be written as follows:

$$r_{t+h} = \theta + \beta_{adj} x_t + \gamma(x_{t+h} - \rho_0 x_{t+h-1}) + \eta_{t+h} \tag{5}$$

where $\beta_{adj} = \beta - \gamma(\rho - \rho_0)$ is the adjusted slope coefficient, which can be simply estimated by applying OLS to Equation (5).

While the presence of persistency and endogeneity in the predictive regression model is a source of bias, the existence of heteroskedasticity is a matter of efficiency. In OLS regression, the variances of errors are always assumed to be constant over time. However, this assumption does not fit well because the stock returns are noisy and heteroskedastic. Westerlund and Narayan (2015a) introduce a FGLS procedure that enables the testing of predictability while, at the same time, also accounting for the information from the heteroskedasticity of the predictor and the regression error. In their set-up, they assume the variance of error follows an autoregressive conditional heteroskedasticity model (ARCH).

The FGLS is also based on Equation (5) but the key difference when compared to the Lewellen AOLS estimator is in the treatment of the persistency of the predictor variable. While

Lewellen (2004) takes the coefficients of the predictor variable ρ to be 0.9999, Westerlund and Narayan (2015a) treat it to be equal to $1+\frac{c}{T}$, where c ≤ 0 is a drift parameter that measures the degree of persistency in x_t. If c =0, then x_t has an exact unit root, and if c <0, x_t is stationary in the sense that ρ approaches one as T increases. Therefore, with FGLS, one does not need to assume that x_t is stationary, as Lewellen (2004) does, which is quite restrictive. Another difference is that FGLS includes the information contained in the ARCH structure of the error terms, which is ignored in the Lewellen AOLS estimator. Therefore, FGLS is expected to outperform AOLS and OLS regression in predicting the stock returns, which is what Westerlund and Narayan (2012) find using a Monte-Carlo simulation exercise.

3.2. Out-of-sample forecasting

Following the earlier studies (Campbell and Thompson, 2008; Rapach *et al*., 2010; Welch and Goyal, 2008), a recursive (expanding) approach is used for out-of-sample forecasting. We divide the total sample size, T, into an in-sample portion composed of the first T_0 observations, and an out-of-sample portion composed of the last, say, P observations. We decide to use several different choices of in-sample periods, including 25%, 50%, and 75% of the total sample. The idea is to test the robustness of the results and, in the process, obviate the issue of data mining. The initial out-of-sample forecasting of the stock returns is as follows:

$$\hat{r}_{i,T_0+1} = \hat{\alpha}_{i,T_0} + \hat{\beta}_{i,T_0} x_{T_0} \tag{6}$$

where $\hat{\alpha}_{i,T_0}$, $\hat{\beta}_{i,T_0}$ are the estimators of α_i and β_i, which can be from the OLS model (α_{OLS}, β_{OLS}); or the AOLS model (α_{AOLS}, β_{AOLS}), or the FGLS model (α_{FGLS}, β_{FGLS}). For the historical average model, $\hat{r}_{T_0+1} = \frac{1}{T_0}\sum_{j=1}^{T_0} r_j$. The first forecasting of the stock returns uses the data up to the first T_0 observations. The forecasted stock returns and all coefficients are re-estimated with new observations over time. By proceeding to the end of the out-of-sample period, we will end up with P forecasts of the stock returns and α, β. For example, the next out-of-sample forecast is:

$$\hat{r}_{i,T_0+2} = \hat{\alpha}_{i,T_{0+1}} + \hat{\beta}_{i,T_{0+1}} x_{T_{0+1}} \tag{7}$$

3.3. Statistical evaluation

The most popular metrics to evaluate the forecast accuracy between models is the MSFE, which has been widely used in the stock return forecasting literature. The MSFE is computed as follows:

$$MSFE = \frac{1}{P}\sum_{j=1}^{P} (r_{T_0+j} - \hat{r}_{T_0+j})^2 \tag{8}$$

where T_0 and P are the number of in-sample and out-of-sample observations, $\hat{r}_{T_0+j}$ is the estimated stock return from the predictive regression model and r_{T_0+j} is the actual stock return. In this paper, we make several forecasting accuracy comparisons between OLS, AOLS, and FGLS (competitor models) and the historical average (benchmark model). We also compare the FGLS against the OLS and AOLS estimators. The well-known out-of-sample R^2 statistic introduced by Campbell and Thompson (2008) is a convenient and simple statistic for comparing the MSFE of the benchmark model ($MSFE_0$) and the MSFE of the competitor model ($MSFE_1$). It is computed as $R_{OS}^2 = 1 -$

$\frac{MSFE_1}{MSFE_0}$. The R_{OS}^2 measures the reduction in MSFE for the competitor model compared to the benchmark model. If the competitor model's MSFE is less than that of the benchmark model ($R_{OS}^2 >$ 0), it indicates that the competitor model is more accurate in forecasting than the benchmark model, and *vice versa*.

Although the R_{OS}^2 statistic can compare the MSFE of forecasting models, there are other relatively traditional tests that allow us to examine the statistical significance of the MSFE from any two models. For example, one can test the null hypothesis $H_0: \ MSFE_0 \ \leq MSFE_1$ against $MSFE_0 > MSFE_1$, corresponding to $H_0: \ R_{OS}^2 \ \leq 0$ against $R_{OS}^2 > 0$. The most popular method in this regard is Diebold and Mariano's (1995) and West's (1996) statistic (DMW), which has the following form:

$$DMW = \sqrt{(T - T_0 - h + 1)}\ \frac{\bar{d}}{\sqrt{\hat{S}}} \tag{9}$$

where

$$\bar{d} = \frac{1}{(T-T_0-h+1)} \sum_{t=T_0}^{T} \hat{d}_{t+h} \tag{10}$$

$$\hat{d}_{t+h} = \hat{u}_{0,t+h}^2 - \hat{u}_{1,t+h}^2 \tag{11}$$

$$\hat{u}_{0,t+h} = r_{t+h} - \hat{r}_{0,t+h} \tag{12}$$

$$\hat{u}_{1,t+h} = r_{t+h} - \hat{r}_{1,t+h} \tag{13}$$

$$\hat{S} = \frac{1}{(T-T_0-h+1)} \sum_{t=T_0}^{T} \left(\hat{d}_{t+h} - \bar{d}\right)^2 \tag{14}$$

where r_{t+h}, $\hat{r}_{0,t+h}$, and $\hat{r}_{1,t+h}$ are the actual excess stock returns, forecasted excess stock returns from benchmark, and competitor models respectively. T and T_0 are the number of observations for entire sample and in-sample periods and *h* is the forecasting horizon. The DMW statistic is equivalent to the t-statistic corresponding to the constant of a regression $\hat{d}_{t+h}$ on a constant, and it has a standard normal asymptotic distribution when compared to non-nested models. However, Clark and McCracken (2001) and McCracken (2007) show that this statistic has a nonstandard distribution when comparing forecasts from nested models. Clark and West (2007) propose a modified Diebold and Mariano (1995) and West (1996) statistic, which they refer to as the MSFE adjusted test. This is obtained by replacing $\hat{d}_{t+h} = \hat{u}_{0,t+h}^2 - \hat{u}_{1,t+h}^2$ with $\tilde{d}_{t+h} = \hat{u}_{0,t+h}^2 - (\hat{u}_{1,t+h}^2 - (\hat{r}_{0,t+h} - \hat{r}_{1,t+h})^2)$. This test statistic is now widely used in the applied time series forecasting literature (for example, Rapach *et al.*, 2010; Kong *et al.*, 2011; and Neely *et al.*, 2011).

3.4. Economic significance

We examine the economic significance available to investors from the forecasting performance of the FGLS. Following previous studies, such as Marquering and Verbeek (2004), Campbell and Thompson (2008), Rapach *et al.* (2010), and Westerlund and Narayan (2012), we analyse the utility gains available for a mean-variance investor. Specifically, we compute the average utility for a mean-variance investor who allocates her portfolio between risky asset and risk-free asset, and who aims to maximise her utility function, which is assumed to be given by:

$$\text{Max} \left[E(r_{t+h}|I_t) - \frac{\gamma}{2} Var(r_{t+h}|I_t)\right] \tag{15}$$

where γ is the relative risk aversion parameter; $E(r_{t+h})$ and $Var(r_{t+h})$ denote the expected values and variance of index excess returns estimated by the forecast approaches. The return on a portfolio of risky asset and risk-free asset is defined as:

$$r_{t+h}^{port} = r_{t+h}^{f} + \omega_{t+h}\, r_{t+h} \qquad (16)$$

where ω_t denotes the proportion of the portfolio allocated to risky assets. The risky asset weight ω_t is positively related to the expected excess return and negatively related to its conditional variance. In other words, an investor will invest more in the risky asset as its return is increasing, and will be equally discouraged from investing if its variance is rising over time. The optimal portfolio weight for risky asset can be obtained as follows:

$$\omega_{t+h}^{*} = \frac{E_t(r_{t+h})}{\gamma Var_t(r_{t+h})} \qquad (17)$$

We first compute the average utility for a mean-variance investor with relative risk aversion parameter,γ, who allocates her portfolio between risky assets and risk-free asset using h-period-ahead forecasts of returns based on the FGLS and historical average models. Following Westerlund and Narayan (2012), we utilise two risk aversion parameters, γ =3 and γ =6. Next, we measure the utility gain as the difference in utility between the FGLS and historical average models, and express the utility gain in the annualised percentage. The utility gain can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in the FGLS predictive regression model. We follow Campbell and Thompson's (2008) study and allow for 50% borrowing at the risk-free rate and no short-selling. Therefore, this restricts the optimal portfolio weight for the risky asset to lie between 0 and 1.50 for each transaction. Furthermore, we allow a transaction cost of 0.1% each time a long or short position is established (Lee and Mathur, 1996a, b; Szakmary and Mathur, 1997; and Narayan *et al.*, 2013).

4. Empirical results

4.1. Preliminary result

The daily, weekly, and monthly prices for the S&P500 index from 4 January 1988 to 31 December 2012 are collected from the Bloomberg database, while the price of WTI crude oil is obtained from the Energy Information Administration[7]. Stock returns are measured as a continuous compounded return and the three-month Treasury bill rate is used to calculate the excess returns. We use different in-sample periods with the proportions 25%, 50%, and 75% of the full sample to forecast the out-of-sample stock returns. As a result, the three out-of-sample periods are April 1994 to December 2012, July 2000 to December 2012, and October 2006 to December 2012.

[7] We also checked whether using a different oil price series will influence our results. To test this, instead of the WTI crude oil price we used the Brent crude oil price in all our models. All results were reproduced. To conserve space, they are not reported here; however, they are available upon request. We find that our results are robust to the use of a different oil price series.

Initially, a number of selective descriptive statistics for the full sample and the three out-of-sample periods are reported in Table 1. Considering the mean value, the stock returns are positive over the full sample period and over the period April 1994 to December 2012, but they become negative over the two recent out-of-sample periods. The mean value of the oil price increases over time and is highest in the period October 2006 to December 2012. On the other hand, oil price volatility has declined over the years, while the volatility of stock returns has increased. The aforementioned features are robust across data frequencies.

Table 1 also reports results for the tests of persistency and heteroskedasticity of the variables. The first-order autoregressive coefficient, ρ, suggests that the oil price is very persistent, as the coefficient is close to one in every period. On the other hand, as expected, the persistency does not exist for the stock returns. This result seems to hold across data frequencies and sample periods. The last two columns under each data frequency refer to the p–value of the Lagrange Multiplier test of the zero slope restriction in an ARCH regression of order q. The null hypothesis of no ARCH effect is easily rejected at the 1% level of significance for both oil price and stock returns across sample periods in the case of the daily and weekly data. A similar result is found for monthly data, although it is slightly weaker. In most cases, the result supports the existence of heteroskedasticity in the data, except we fail to reject the null of no ARCH over the out-of-sample period October 2006 - December 2012 in the case of monthly frequency.

INSERT TABLE 1

Table 2 presents the results for the endogeneity test of variables by two instruments. First, we compute the correlation and its p-value between the OLS regression residuals in Equation (1) and the first difference of the predictor variable. The correlations are positive and vary in the 0.154 to 0.322 range for daily data. For weekly and monthly data frequencies, correlations fall in the range 0.184 to 0.370, and 0.220 to 0.512, respectively. All correlations are statistically significant at the 1% level across all three data frequencies. Secondly, we compute γ, which is the slope coefficient from Equation (3), by regressing residuals ϵ_t from predictive regression (1) on residuals ε_t from the autoregressive regression (2). The OLS estimator assumes that these residuals are uncorrelated to each other (which means γ = 0), but γ is significantly different from 0 at the 5% level at least, suggesting that endogeneity does exist. We, therefore, conclude that oil price is endogenous.

INSERT TABLE 2

To conclude, the preliminary results strongly suggest the existence of the persistency and endogeneity of the predictor variable, and heteroskedasticity in the predictive regression model. Therefore, the FGLS approach is expected to outperform the other models in forecasting the equity premium, as discussed earlier.

4.2. Out-of-sample result

We employ three different predictive regression estimators, namely, the OLS, AOLS, and FGLS to forecast the stock returns of the S&P500 index using the oil price as a predictor. As mentioned in the previous section, different initial in-sample periods are considered to compute out-of-sample forecasts through employing a recursive estimation window. The out-of-sample forecasting accuracy is evaluated by the well-known R^2_{OS} statistic, which indicates that the competitor forecasting model outperforms the benchmark model when R^2_{OS} takes a positive value, and *vice versa*. To test for the statistical significance of the outperformance, the DMW statistic is used for non-nested model comparisons, and the MSFE-adjusted statistic is used for nested model comparisons. The forecasting evaluation is undertaken for the four horizons, and the sign restriction, which is believed to improve the forecasting ability of models, is also applied.

4.2.1. Comparison between OLS, AOLS and FGLS

Table 3 reports results which compare the performance of the FGLS to OLS and AOLS approaches based on the daily (Panel A), weekly (Panel B), and monthly (Panel C) data frequencies. The table presents R^2_{OS} and p-value for the DMW test statistic, with the null hypothesis that $R^2_{OS} \leq 0$ against the alternative that $R^2_{OS} > 0$. Starting with the daily data, we observe that all of the R^2_{OS} statistics are positive, suggesting that using FGLS instead of OLS or AOLS can improve the accuracy of forecasting the stock returns. The magnitudes of R^2_{OS} are in the range 0.068% to 0.104% and 0.068% to 0.120% for OLS and AOLS, respectively. Regarding the statistical significance of the FGLS outperformance, we see that the null hypothesis that $R^2_{OS} \leq 0$ is rejected at the 10% level of significance or higher in most cases for both OLS and AOLS-based results. These results are consistent throughout different in-sample sizes and forecasting horizons. Similar results are obtained when we use weekly data. The superior performance of the FGLS estimator still holds up to the 12-week forecasting horizon. Again, all the R^2_{OS} are positive indicating that FGLS is better than the OLS model, and 67% (83%) of them are statistically significant at the 10% level or higher in OLS (AOLS) based results. On the other hand, the results still favour the FGLS in the case of monthly data, although the evidence is not as strong when models are estimated using data at daily and weekly frequency. We find 25% of the R^2_{OS} is statistically significant at the 10% level in both OLS and AOLS-based results.

In summary, three main features of the results are observed. First, the FGLS model outperforms the OLS and AOLS models in forecasting stock return. Second, the outperformance is statistically significant. Finally, the results are robust across the four different forecasting horizons and three different in-sample sizes used in this study, but the same cannot be said in regard to the use of different data frequencies. The superiority of the FGLS model is stronger in the case of the daily data compared to the results based on the weekly and monthly frequencies.

INSERT TABLE 3

4.2.2. Comparison to historical average benchmark model

The historical average model is widely used in the stock return forecasting literature as a benchmark model (Welch and Goyal, 2008; Campbell and Thompson, 2008; Rapach *et al.*, 2010). Therefore, we also compare the OLS, AOLS, and FGLS predictive regression models based on oil price as a predictor with the historical average model, starting with Table 4, which presents the results of the comparison between the OLS and AOLS predictive regression models to the historical average model. The table reports the R_{OS}^2 statistic and the p-value of the MSFE-adjusted statistic for assessing the statistical significance of the corresponding forecasts under the null hypothesis that the competitor forecasts (OLS, AOLS) are not better than the benchmark forecasts (historical average). Focusing on the OLS estimator results, we find that the OLS estimator cannot beat the historical average model, since the R_{OS}^2 statistics are negative in all cases. We also observe similar results for the comparison between the AOLS and the historical average models. In addition, the results are consistent across the choices of in-sample periods and forecasting horizons. We conclude that the OLS and AOLS based oil price estimators are not superior to the historical average model in forecasting the stock returns.

INSERT TABLE 4

Next, we compare the FGLS and restricted FGLS models which apply the sign restriction of Campbell and Thompson (2008)[8], with the historical average model and report the results in Table 5. Three observations are worth highlighting. First, we find that the results are in favour of FGLS and restricted FGLS models over the historical average model in forecasting stock returns. In addition, the superiority of these two models is statistically significant. Second, we also observe that the results are data frequency-dependent. Finally, the sign restriction can improve the stock return forecasting accuracy.

Focusing on the FGLS model, the results based on daily and weekly data frequencies strongly suggest that the FGLS estimator is better than the historical average benchmark model, as the R_{OS}^2 takes positive values in all cases across different in-sample periods and forecasting horizons. The R_{OS}^2 ranges from 0.019% to 0.095%, and from 0.042% to 0.795% in the case of daily and weekly data frequencies, respectively. When we consider daily data, R_{OS}^2 is statistically significantly greater than zero at the 10-day forecasting horizon when we use 25% and 50% in-sample periods. In the case of weekly data frequency, R_{OS}^2 is statistically significantly greater than zero at the first two forecasting horizons in all three in-sample periods, and also when we consider the 12-week forecasting horizon with a 75% in-sample period. On the other hand, the results are mixed for the first two forecasting horizons when we used monthly data, as we obtain both positive and negative R_{OS}^2. For the forecasting horizons of six-months and 12-months, the results are in favour of the historical average model.

Turning to the restricted FGLS model, it is clearly depicted that the results are better than those in the FGLS model. The improvement is illustrated by the number of statistically significant and greater than zero R_{OS}^2 statistics increasing from two to eight cases when we use daily data. However, this is not obvious based on the weekly data results, but the improvement is very strong in the case of the monthly data. We find six cases where R_{OS}^2 is negative in the FGLS model and results become

[8]Campbell and Thompson (2008) restrict the sign of the predictor coefficients and forecasted returns, which implies the predictor coefficients must have the theoretically expected sign and the forecasted returns must be positive or equal to zero otherwise. In this paper, we only apply the sign restriction for the forecasted returns but not the predictor coefficients, as the literature provides inconclusive results for the effect of crude oil on the equity market.

positive after applying the sign restriction. In addition, we also observe that an additional two R^2_{OS} are statistically significantly and greater than zero.

INSERT TABLE 5

4.3. Economic significance

The utility gains using the FGLS model instead of the historical average to forecast the stock returns are reported in Table 6[9]. From previous sections, we find that the FGLS model provides a more accurate forecasting of the stock returns than the historical average model. Hence, the utility gains are expected to be positive and the results clearly support this expectation. We compute utility gains at risk aversion factors of three and six, and we also report results at all four different forecasting horizons, as defined in the previous section.

We find that utility gains at all forecasting horizons and using both three and six risk aversion factors are positive for daily and weekly data frequencies. In the case of monthly frequency, the utility gains are positive when we use a risk aversion factor equal to three. Our results suggest that mean-variance investors are able to obtain utility gains by using the FGLS instead of the historical average model. The results are robust across the in-sample periods, forecasting horizons, and risk aversion parameters. Another noteworthy point is that the positive utility gains are quite sizeable. Considering the daily data results, the utility gains vary in a range from 1.294% to 5.196%, and the average value is 3.2%. This result can be interpreted as an investor being willing to pay an extra 3.2% per annum to have access to the additional information available in the FGLS predictive approach. The results are similar based on weekly and monthly data models, where the average utility gains over in-sample periods, forecasting horizons, and risk aversions are 3.4% and 2.6%, respectively.

INSERT TABLE 6

4.4. Sector level analysis

In this section, we use the daily data for ten US sectors categorised by the Global Industry Classification Standard (GICS) from the DataStream database. We focus on the FGLS approach, which is shown in previous sections to be superior to the OLS and AOLS approaches, for forecasting stock returns. We also employ the sign restriction approach of Campbell and Thompson (2008) to forecast stock returns as it can improve the forecasting accuracy. The results based on the comparison between the restricted FGLS model and the historical average model are reported in Table 7. Again, the sectoral results are computed using 25%, 50%, and 75% in-sample periods and four forecasting horizons. Overall, we find strong evidence supporting the outperformance of the restricted FGLS over the historical average model. Out of ten sectors, six have a positive value of R^2_{OS} in most or all of the four forecasting horizons. They are the industrial, consumer staples, consumer discretion, telecommunication, financial, and technology sectors. In addition, the number of rejections of the null hypothesis that $R^2_{OS} \leq 0$ is remarkably high, which means that R^2_{OS} is statistically significantly greater than 0. The results are robust across in-sample sizes and forecasting horizons. For the other four sectors, namely, energy, healthcare, material, and utility, the R^2_{OS} statistics are negative in most

[9] Due to the restriction of no short-selling for risky asset ($\omega_t > 0$), the economic significance results for FGLS and restricted FGLS are the same.

cases, suggesting that the FGLS model cannot beat the historical average model in forecasting stock returns in these sectors

INSERT TABLE 7

The economic significance results based on the restricted FGLS model are reported in Table 8. Results based on a 25% in-sample period (see Panel A) reveal that in the case of seven sectors, namely, industrial, consumer staples, healthcare, consumer discretion, telecom, financial, and technology, the average utility gains are positive, which implies that using the FGLS instead of the historical average forecasting model can bring gains to a mean-variance investor. In addition, the magnitude of the gains is sizeable. On average, the positive utility gains of sectors are heterogeneous, ranging from 0.1% to 2.5% per annum. On the other hand, we find negative utility gains for three sectors (energy, material and utility) which do not have a positive R^2_{OS} , as reported in Table 7. Although it is shown that the FGLS model is not superior to the historical average model in the healthcare sector, it has a positive utility gain of 0.1% per annum on average. The results are consistent across risk aversion parameters and forecasting horizons. Turning to the results in Panels B and C, we find similar features. There are seven sectors associated with positive average utility gains and their values range from 0.7% to 3.8% and 0.4% to 9% per annum when we consider in-sample sizes of 50% and 75%, respectively.

INSERT TABLE 8

4.5. Summary of key findings

This sub-section is devoted to summarising the key findings from our forecasting exercise. Essentially this analysis offers four key messages which we highlight here. The first message is that oil prices are a useful predictor of stock returns. The second message is that the ability of oil prices to predict stock returns out-of-sample is data frequency dependent. The third message relates to the choice of the forecasting estimator. Granted that oil price is a useful predictor of stock returns, one must not ignore the choice of estimator on hand. We show that an estimator, such as FGLS, which caters of the key salient features of the data—namely, heteroskedasticity and predictor persistency and endogeneity—works best when the data on hand is of a relatively high frequency nature, such as daily and weekly.

Our final message is that when we extend the analysis to forecasting sectoral stock returns, there are some sectors (energy, material, and utility) for which a constant returns model actually beats the oil price-based forecasting model. This implies that for these sectors the information content in oil price is not sufficient to forecast sectoral returns better than a simple constant returns model. One possible reason for this outcome could be that these three sectors are oil-related sectors in terms of consumption (materials and utility) and production (energy) where oil price is not a surprise indicator. In other words, for these sectors the effect of oil price is expected.

4.6. Determinants of sector return predictability

The results reported in Table 7 provide strong evidence of the heterogeneity in stock return forecasting among sectors. Using daily data, the crude oil price is able to predict stock returns in a number of sectors but not for every sector. In addition, the magnitude and the number of statistically significant R^2_{OS} statistics also vary from sector to sector. These interesting results raise a new question that is worthy of investigation: what determines the predictability of sector stock returns? In order to answer this question, we investigate the relationship between the predictability and the sectoral characteristic variables. The proxy for predictability of each sector is the difference between the forecasting errors from the historical average and FGLS models. It is computed as follows:

$$d_{t+1} = (r_{t+1} - \hat{r}_{0,t+1})^2 - (r_{t+1} - \hat{r}_{1,t+1})^2 \tag{18}$$

where $\hat{r}_{0,t+1}$ is the forecasted stock return of the historical average model and $\hat{r}_{1,t+1}$ is from the FGLS model. The higher value of this proxy implies that the predictability of stock returns using the crude oil price is stronger. The time series for the predictability variables is generated from out-of-sample forecasting based on 25%, 50% and 75% in-sample periods.

The sectoral characteristic variables for this analysis are the book-to-market ratio (BM), dividend yield (DY), price-earnings ratio (PE), trading volume (TV), and market capitalisation (MV). All independent variables are taken in natural logarithmic form. Unit root tests reveal that except for the predictability variable, all other variables are non-stationary. Therefore, we take the first difference of these variables and fit them in a GARCH (1,1) model.

$$d_t = \alpha + \ +\beta_{BM}BM_t + \beta_{DY}DY_t + \beta_{PE}PE_t + \beta_{TV}TV_t + \beta_{MV}MV_t \ + \varepsilon_t$$
$$\sigma_t^2 = \gamma_0 + \gamma_1\varepsilon_{t-1}^2 + \gamma_2\sigma_{t-1}^2 \tag{19}$$

where d_t, the proxy for predictability of each sector, is the difference between the forecasting errors from the historical average and FGLS models, computed as Equation (18). The results are reported in Table 9. The estimated coefficients together with their p-values (in parentheses) are reported. A number of noteworthy features from the results are as follows. First, the book-to-market ratio, dividend yield, and trading volume have a positive effect on the predictability of sectoral stock returns. These results can be interpreted as the sectors that have a higher book-to-market ratio, dividend yield, and trading volume which have stronger stock return predictability using the crude oil price. On the other hand, the sector stock return predictability is negatively impacted by the size and price-earnings ratio. The results of the book-to-market ratio and the size are consistent with Kong *et al.* (2011), who examine 25 Fama-French size and value-sorted portfolios and find that predictability is stronger for the portfolios with low market capitalisation.

Secondly, the impact of sectoral characteristics, especially when we consider book-to-market, dividend yield, and size variables, on predictability of sectoral stock returns are statistically significant. The results are slightly weaker in the case of price-earnings ratio and trading volume but the proportion of rejected null hypotheses is still high, from 50% to 80% of the cases. Thirdly, the impact of the characteristic variables (where they are statistically significant) varies from sector-to-sector. For example, the coefficients of the book-to-market ratio range from -0.028 to 1.104, implying that the book-to-market ratio explains the predictability of stock returns in some sectors being stronger than in others. Finally, the results are robust across all three in-sample sizes of 25%, 50% and 75%.

INSERT TABLE 9

5. Conclusions

This paper uses the crude oil price to predict stock returns using S&P500 indices. We apply a recently introduced FGLS forecasting model, which takes into account three potential issues in a predictive regression, namely, heteroskedasticity, predictor endogeneity, and persistency. We use daily, weekly, and monthly data over the period 4 January 1988 to 31 December 2012. Three choices of in-sample periods with the proportions 25%, 50% and 75% of the full sample are utilised to forecast the out-of-sample stock returns. The empirical analyses are reproduced with a different oil price series to test the robustness of the results. The main findings and contributions of this paper are as follows.

First, unlike the extant literature on oil price and stock returns, we focus on out-of-sample forecasting of returns. We show that how well the oil price forecasts stock returns depends not only on the data frequency but also on the estimator. We employ pair-wise comparisons between the OLS, AOLS and FGLS models and find that the FGLS model is superior to the others, as expected. Compared to the historical average benchmark model, the FGLS model using the crude oil price as a predictor performs better than the historical average model in forecasting stock returns, but it is opposite for the OLS and AOLS models .The outperformance of the FGLS over the other models is data frequency-dependent. The outperformance of FGLS is strongly evidenced in the daily and weekly data frequencies and it is found to be weaker when we use monthly data. In addition, the results from this paper also support the findings of Campbell and Thompson (2008) that a sign restriction can improve the forecasting power. Turning to the economic significance, the results strongly suggest that a mean-variance investor would be better off in terms of utility using the FGLS model rather than the historical average model.

Second, out-of-sample forecasting of returns is sector-dependent suggesting that the oil price is relatively more important for some sectors than others. We find strong evidence of the superiority of the FGLS model while predicting stock returns using the crude oil price in six out of ten sectors, while the results based on the other four sectors (namely, energy, healthcare, material, and utility) are in favour of the historical average model. In addition, the results are also heterogeneous among sectors in terms of economic significance.

Third, we examine the determinants of predictability for each sector using industry characteristics, and find strong evidence that return predictability has links to certain industry characteristics, such as book-to-market ratio, dividend yield, size, price-earnings ratio, and trading volume. Furthermore, we find that the book-to-market ratio, dividend yield, and trading volume have a positive effect, while the size and price-earnings ratio have a negative impact on sector stock return predictability.

Reference

Amemiya T, Wu R. 1972. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* **67**: 628–632.

Andersen T, Bollerslev T, Lange S. 1999. Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance* **6**: 457-477.

Ang A, Bekaert G. 2007. Return predictability: is it there? *Review of Financial Studies* **20**: 651-707.

Avramov D, Chordia T. 2006. Predicting stock returns. *Journal of Financial Economics* **82**: 387-415.

Ball R. 1978. Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics* **6**: 103–126.

Banz R. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* **9**: 3-18.

Bossaerts P, Hillion P. 1999. Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies* **12**: 405-428.

Boswijk H, Klaassen F. 2012. Why frequency matters for unit root testing in financial time series. *Journal of Business and Economic Statistics* **30**: 351-357.

Boudoukh J, Richardson M, Whitelaw R. 2008. The myth of long-horizon predictability. *Review of Financial Studies* **21**: 1577–1605.

Brennan M, Xia Y. 2005. Tay's as good as cay. *Finance Research Letter* **2**: 1-14.

Butler A, Grullon G, Weston J. 2005. Can managers forecast aggregate market return? *Journal of Finance* **60**: 963-986.

Campbell J. 1987. Stock returns and the term structure. *Journal of Financial Economics* **18**: 373-399.

Campbell J, Shiller R. 1988a. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* **1**: 195-228.

Campbell J, Shiller R. 1988b. Stock prices, earnings and expected dividends. *Journal of Finance* **43:** 661-676.

Campbell J, Thompson S. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* **24**: 1509-1531.

Chen G, Firth M, Rui O. 2001. The dynamic relationship between stock returns, trading volume, and volatility. *The Financial Review* **36**: 153- 174.

Chen N, Roll R, Ross S. 1986. Economic forces and the stock market. *Journal of Business* **59**: 383–403.

Clark P. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41**: 135–155.

Clark T, McCracken M. 2001. Test of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* **105**: 85-110.

Clark T, West K. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* **138**: 291-311.

Connor G, Korajczyk R. 1989. An intertemporal equilibrium asset pricing model. *Review of Financial Studies* **2**: 373-392.

Copeland T. 1976. A model of asset trading under the assumption of sequential information arrival. *Journal of Finance* **31**: 1149–1168.

Diebold F, Mariano R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253-263.

Driesprong G, Jacobsen B, Maat B. 2008. Striking oil: Another puzzle? *Journal of Financial Economics* **89**: 307-327.

Elton E, Gruber M, Blake C, Krasny Y, Ozelge S. 2010. The effect of holding data frequency on conclusions about mutual fund behaviour. *Journal of Banking and Finance* **34**: 912-922.

Fama E. 1981. Stock returns, real activity, inflation and money. *American Economic Review* **71**: 545–565.

Fama E, French K. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22: 3-25.

Fama E, French K. 1989. Business conditions and expected return on stocks and bonds. *Journal of Financial Economics* 25: 23-49.

Fama E, French K. 1991. The cross-section of expected stock returns. *Journal of Finance* 47: 427-465.

Ferson W, Harvey C. 1991. The variation in economic risk premiums. *Journal of Political Economy* **99**: 385-415.

Foster A. 1995. Volume–volatility relationships for crude oil futures. *Journal of Futures Markets* **15**: 929–951.

Gjerde O, Saettem F. 1999. Causal relations among stock returns and macroeconomic variables in a small, open economy. *Journal of International Financial Market, Institutions and Money* **9**: 61-74.

Goyal A, 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**: 228-248.

Hooker M. 1993. Testing for cointegration: Power versus frequency of observation. *Economics Letters* **41**: 359-363.

Huang R, Masulis R, Stoll H. 1996. Energy shocks and financial markets. *Journal of Futures Markets* **16**: 1–27.

Jones C, Kaul G. 1996. Oil and the stock market. *Journal of Finance* **51**: 463- 491.

Jones C, Kaul G, Lipson M. 1994. Transactions, volume and volatility. *Review of Financial Studies* **7**: 631–651.

Karpoff J. 1987. The relationship between price changes and trading volume: a survey. *Journal of Financial and Quantitative Analysis* **22**: 109-126.

Kirby C. 1998. The restrictions on predictability implied by rational asset pricing models. *Review of Financial Studies* **11**: 343-382.

Kong A, Rapach D, Strauss J, Zhou G. 2011. Predicting market components out of sample: asset allocation implications. *Journal of Portfolio Management* **37**: 29-41.

Kothari S, Shanken J. 1997. Book-to-market, dividend yield and expected market returns: a time series analysis. *Journal of Financial Economics* **44**: 169-203.

Lahiri K, Mamingi N. 1995. Testing for cointegration: Power versus frequency of observation—another view. *Economics Letters* **49**: 121-124.

Lamont O. 1998. Earnings and expected returns. *Journal of Finance* **53**: 1563-1587.

Lee C, Mathur I. 1996a. Trading rule profits in European currency spot cross-rates. *Journal of Banking and Finance* **20**: 949-962.

Lee C, Mathur I. 1996b. A comprehensive look at the efficacy of technical trading rules applied to cross-rates. *European Journal of Finance* **2**: 389-411.

Lettau M, Nieuwerburgh S. 2008. Reconciling the return predictability evidence. *Review of Financial Studies* **21**: 1607–1652.

Lewellen J. 2004. Predicting returns with financial ratios. *Journal of Financial Economics* **74**: 209-235.

Llorente G, Michaely R, Saar G, Wang J. 2002. Dynamic volume-return relation of individual stocks. *Journal of Financial Studies* **15**: 1005–1047.

Marquering W, Verbeek M. 2004. The economic value of predicting stock index returns and volatility. *The Journal of Financial and Quantitative Analysis* 39: 407-429.

McCracken M. 2007. Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics* **140**: 719-752.

Miller J, Ratti R. 2009. Crude oil and stock markets: Stability, instability and bubbles. *Energy Economics* **31**: 559-568.

Narayan P, Narayan S, Sharma S. 2013. An analysis of commodity markets: What gain for investors? *Journal of Banking and Finance* **37**: 3878-3889.

Narayan P, Narayan S. 2010. Modelling the impact of oil prices on Vietnam's stock prices. *Applied Energy* **87**: 356-361.

Narayan P, Sharma S. 2011. New evidence on oil price and stock returns. *Journal of Banking and Finance* **35**: 3253-3262.

Narayan P, Narayan S, Thuraisamy K. 2014a. Can institutions and macroeconomic factors predict stock returns in emerging markets? *Emerging Markets Review* **19**: 77-95.

Narayan P, Narayan S, Prabheesh P. 2014b. Stock returns, mutual fund flows and spillover shocks. *Pacific-Basin Finance Journal* **29**: 146-162.

Narayan P, Sharma S, Poon W, Westerlund J. 2014c. Do oil prices predict economic growth? New global evidence. *Energy Economics* **41**: 137-146.

Narayan P, Amhed H, Narayan S. 2015a. Do momentum-based trading strategies work in the commodity futures markets? *Journal of Futures Markets*, DOI: 10.1002/fut.21685.

Narayan P, Sharma S, Thuraisamy K. 2015b. Can governance quality predict stock market returns: New global evidence? *Pacific-Basin Finance Journal* (In Press).

Narayan P, Sharma S. 2015. Does data frequency matter for the impact of forward premium on the spot exchange rate? *International Review of Financial Analysis* (In Press).

Nelson C, Kim M. 1993. Predictable stock return: the roles of small sample bias. *Journal of Finance* **48**: 641-661.

Nijman T, Palm F. 1990a. Parameter identification in ARMA-processes in the presence of regular but incomplete sampling. *Journal of Time Series Analysis* **11**: 239–248.

Nijman T, Palm F. 1990b. Predictive accuracy gain from disaggregate sampling in ARIMA models. *Journal of Business and Economic Statistics* **8**: 405–415.

Otero J, Smith J. 2000. Testing for cointegration: Power versus frequency of observations—Further Monte Carlo results. *Economics Letters* **67**: 5-9.

Park J, Ratti R. 2008. Oil price shocks and stock markets in the U.S. and 13 European countries. *Energy Economics* **30**: 2587-2608.

Pontiff J, Schall L. 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics* **49**: 141-160.

Rapach D, Strauss J, Tu J, Zhou G. 2011. Out-of-Sample industry return predictability: Evidence from a large number of predictors. Working paper, Saint Louis University.

Rapach D, Strauss J, Zhou G. 2010. Out-of-sample equity premium prediction: Combination forecast and links to the real economy. *Review of Financial Studies* **23**: 821-862.

Rapach D, Wohar M, Rangvid J. 2005. Macro variables and international stock return predictability. *International Journal of Forecasting* **21**: 137-166.

Rozeff M. 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management* **11**: 68–75.

Sadorsky P. 1999. Oil price shocks and stock market activity. *Energy Economics* **21**: 449-469.

Stambaugh R. 1983. Arbitrage pricing with information. *Journal of Financial Economics* **12**: 357-370.

Stambaugh R. 1999. Predictive regressions. *Journal of Financial Economics* **54**: 375-421.

Szakmary A, Mathur I. 1997. Central bank intervention and trading rule profits in foreign exchange markets. *Journal of International Money and Finance* **16**: 513–535.

Tashman L. 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* **16**: 437-450.

Tauchen G, Pitts M. 1983. The price variability-volume relationship on speculative markets. *Econometrica* **51**: 485-505.

Wang G, Yau J. 2000. Trading volume, bid-ask spread and price volatility in futures markets. *Journal of Futures Markets* **20**: 943–970.

Wei C. 2003. Energy, the stock market and the Putty-Clay investment model. *American Economic Review* **93**: 311–23.

Welch I, Goyal A. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21: 1455-1508.

West K. 1996. Asymptotic inference about predictive ability. *Econometrica* 64: 1067-1084.

Westerlund J, Narayan P. 2012. Does the choice of estimator matter when forecasting stock returns. *Journal of Banking and Finance* **36**: 2632-2640.

Westerlund J, Narayan P. 2015a. Testing for predictability in conditionally heteroskedastic stock returns. *Journal of Financial Econometrics*, DOI: 10.1093/jjfinec/nbu001.

Westerlund J, Narayan P. 2015b. Testing for predictability in conditionally heteroskedastic stock returns. *Journal of Financial Econometrics*, http://dx.doi.org/10.1093/jjfinec/nbu003.

Table 1 : Selective descriptive statistics

	Daily data					Weekly data					Monthly data				
	Mean	SD	ρ	ARCH(5)	ARCH(10)	Mean	SD	ρ	ARCH(6)	ARCH(12)	Mean	SD	ρ	ARCH(6)	ARCH(12)
Jan 1988 - December 2012															
Oil price	40.488	28.860	1.000	0.00	0.00	40.541	28.891	0.997	0.00	0.00	40.849	29.201	0.991	0.00	0.00
S&P500	0.013	1.157	-0.057	0.00	0.00	0.064	2.346	-0.094	0.00	0.00	0.272	4.292	0.057	0.00	0.00
April 1994 – December 2012															
Oil price	47.343	30.280	0.999	0.00	0.00	47.430	30.294	0.996	0.00	0.00	47.864	30.603	0.987	0.00	0.00
S&P500	0.012	1.244	-0.068	0.00	0.00	0.060	2.512	-0.082	0.00	0.00	0.267	4.520	0.097	0.00	0.07
July 2000 - December 2012															
Oil price	61.230	28.077	0.999	0.00	0.00	61.348	28.098	0.993	0.00	0.00	61.988	28.326	0.973	0.00	0.03
S&P500	-0.009	1.341	-0.088	0.00	0.00	-0.046	2.650	-0.046	0.00	0.00	-0.163	4.686	0.167	0.01	0.02
October 2006 - December 2012															
Oil price	82.771	20.251	0.995	0.00	0.00	82.886	20.332	0.978	0.00	0.00	84.072	19.603	0.901	0.39	0.80
S&P500	-0.001	1.540	-0.120	0.00	0.00	-0.011	3.028	-0.060	0.00	0.00	-0.048	5.170	0.248	0.27	0.68

Notes: This table reports the selective descriptive statistics for the WTI crude oil price and the S&P500 index excess returns. Stock returns are measured as a continuous compounded return and the three-month Treasury bill rate is used to calculate the excess return. The descriptive statistics are for the total sample from 4 January 1988 to 31 December 2012 and three out-of-sample periods (associated with three choices of in-sample period, 25%, 50% and 75%). The table reports results for the daily, weekly and monthly data. ρ refers to the autoregressive coefficient in Equation (2) $x_{t+1} = \mu(1-\rho) + \rho x_t + \varepsilon_{t+1}$, ARCH (q) refers to a Lagrange multiplier test of the zero slope restriction in an ARCH regression of order q, and the p-value of the test is reported.

Table 2: Endogeneity test

	Daily data		Weekly data		Monthly data	
	Correlation	γ	Correlation	γ	Correlation	γ
January 1988 – December 2012	0.154	0.152	0.184	0.178	0.220	0.199
	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)
April 1994 – December 2012	0.194	0.186	0.222	0.205	0.297	0.253
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
July 2000 - December 2012	0.228	0.196	0.262	0.214	0.377	0.280
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
October 2006 - December 2012	0.322	0.250	0.370	0.272	0.512	0.329
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Notes: This table reports the endogeneity test in the total sample from 4 January 1988 to 31 December 2012 and three out-of-sample periods (associated with three choices of in-sample period, 25%, 50% and 75%). The first column of each data frequency panel presents the correlation between OLS regression residuals in Equation (1) $r_{t+h} = \alpha + \beta x_t + \epsilon_{t+h}$ and first difference of predictor (crude oil price). γ denotes the coefficient in equation (3) $\epsilon_t = \gamma \varepsilon_t + \eta_t$. The p-value with the null hypotheses that the correlation is equal to zero and the $\gamma = 0$ is reported in parentheses.

Table 3: FGLS versus OLS and AOLS

Panel A : Daily data									
		h =1		**h =2**		**h =5**		**h =10**	
OLS	25%	0.073**	(0.02)	0.068***	(0.01)	0.063*	(0.07)	0.079***	(0.01)
	50%	0.089**	(0.02)	0.082***	(0.01)	0.083*	(0.06)	0.101***	(0.01)
	75%	0.100**	(0.05)	0.086**	(0.04)	0.082	(0.13)	0.104**	(0.05)
AOLS	25%	0.078***	(0.01)	0.068***	(0.01)	0.065**	(0.04)	0.079***	(0.01)
	50%	0.098***	(0.01)	0.084***	(0.01)	0.089**	(0.03)	0.103***	(0.01)
	75%	0.120**	(0.03)	0.097**	(0.03)	0.088*	(0.10)	0.105**	(0.04)
Panel B : Weekly data									
		h =1		**h =2**		**h =4**		**h =12**	
OLS	25%	0.460*	(0.08)	0.557*	(0.08)	0.389*	(0.10)	0.492*	(0.09)
	50%	0.601*	(0.08)	0.762*	(0.07)	0.528*	(0.10)	0.581	(0.12)
	75%	0.804*	(0.08)	0.861	(0.11)	0.651	(0.13)	0.801	(0.12)
AOLS	25%	0.471*	(0.06)	0.571*	(0.06)	0.390	(0.11)	0.892***	(0.01)
	50%	0.615*	(0.06)	0.780**	(0.05)	0.572*	(0.06)	1.273***	(0.01)
	75%	0.787*	(0.08)	0.844	(0.11)	0.741*	(0.09)	1.719**	(0.02)
Panel C : Monthly data									
		h =1		**h =3**		**h =6**		**h =12**	
OLS	25%	0.492	(0.23)	0.298	(0.37)	-0.016	(0.51)	2.088**	(0.03)
	50%	0.490	(0.27)	0.624	(0.30)	0.046	(0.49)	3.375***	(0.01)
	75%	1.298	(0.13)	2.455**	(0.05)	0.525	(0.39)	2.461	(0.11)
AOLS	25%	0.451	(0.28)	0.872	(0.34)	2.962	(0.05)	3.303**	(0.03)
	50%	0.422	(0.27)	2.406	(0.19)	3.421	(0.06)	4.925***	(0.01)
	75%	1.447**	(0.05)	5.895***	(0.01)	4.537	(0.01)	1.140	(0.29)

Notes: This table reports the R^2_{OS} for the competitor model FGLS compared to the benchmark models OLS and AOLS. The in-sample proportion choices are in the second column. The DMW p-value in parentheses tests the null hypothesis $H_0: \; MSFE_0 \leq MSFE_1$ against $MSFE_0 > MSFE_1$, corresponding to $H_0: \; R^2_{OS} \leq 0$ against $R^2_{OS} > 0$. *h* refers to the forecasting horizon. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively

Table 4: OLS and AOLS versus historical average

				Panel A : Daily data					
		h =1		**h =2**		**h =5**		**h =10**	
OLS	25%	-0.054	(0.28)	-0.047	(0.28)	-0.035	(0.23)	-0.030	(0.22)
	50%	-0.047	(0.23)	-0.036	(0.23)	-0.032	(0.20)	-0.029	(0.20)
	75%	-0.026	(0.31)	-0.015	(0.31)	-0.001	(0.27)	-0.009	(0.29)
AOLS	25%	-0.061	(0.28)	-0.048	(0.26)	-0.041	(0.21)	-0.033	(0.21)
	50%	-0.056	(0.23)	-0.038	(0.21)	-0.038	(0.20)	-0.030	(0.19)
	75%	-0.047	(0.35)	-0.026	(0.35)	-0.007	(0.30)	-0.011	(0.31)
				Panel B : Weekly data					
		h =1		**h =2**		**h =4**		**h =12**	
OLS	25%	-0.243	(0.20)	-0.220	(0.18)	-0.348	(0.35)	-0.443	(0.45)
	50%	-0.237	(0.17)	-0.257	(0.18)	-0.354	(0.30)	-0.363	(0.37)
	75%	-0.086	(0.27)	-0.066	(0.26)	-0.235	(0.36)	-0.177	(0.29)
AOLS	25%	-0.255	(0.20)	-0.234	(0.18)	-0.349	(0.33)	-0.849	(0.78)
	50%	-0.251	(0.17)	-0.275	(0.18)	-0.398	(0.30)	-1.067	(0.75)

	75%	-0.069	(0.29)	-0.049	(0.28)	-0.326	(0.45)	-1.113	(0.77)

Panel C : Monthly data									
		h =1		**h =3**		**h =6**		**h =12**	
OLS	25%	-1.188	(0.25)	-2.723	(0.47)	-3.150	(0.58)	-5.347	(0.81)
	50%	-1.000	(0.19)	-2.473	(0.40)	-4.162	(0.57)	-6.997	(0.77)
	75%	-0.280	(0.26)	-0.767	(0.25)	-5.009	(0.52)	-4.911	(0.67)
AOLS	25%	-1.146	(0.20)	-3.318	(0.58)	-6.315	(0.86)	-6.671	(0.34)
	50%	-0.931	(0.16)	-4.345	(0.56)	-7.803	(0.83)	-8.740	(0.36)
	75%	-0.431	(0.30)	-4.451	(0.63)	-9.423	(0.94)	-3.509	(0.65)

Notes: This table reports the R^2_{OS} for the competitor models, OLS and AOLS, compared with the historical average benchmark model. The in-sample proportion choices are in the second column. The Clark and West (2007) adjusted MSFE p-values are reported in parentheses, which test the null hypothesis $H_0: MSFE_0 \leq MSFE_1$ against $MSFE_0 > MSFE_1$, corresponding to $H_0: R^2_{OS} \leq 0$ against $R^2_{OS} > 0$. *h* refers to the forecasting horizon. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 5: FGLS and restricted FGLS versus historical average

Panel A : Daily data									
		h =1		**h =2**		**h =5**		**h =10**	
FGLS	25%	0.019	(0.15)	0.021	(0.14)	0.028	(0.13)	0.049*	(0.10)
	50%	0.043	(0.12)	0.046	(0.11)	0.051	(0.11)	0.072*	(0.08)
	75%	0.074	(0.18)	0.072	(0.18)	0.082	(0.18)	0.095	(0.15)

Restricted	25%	0.037**	(0.05)	0.036*	(0.06)	0.047**	(0.04)	0.042**	(0.05)
FGLS	50%	0.066**	(0.02)	0.064**	(0.02)	0.070***	(0.01)	0.062**	(0.03)
	75%	0.022	(0.20)	0.020	(0.23)	0.024	(0.20)	0.018	(0.25)
Panel B : Weekly data									
		h =1		**h =2**		**h =4**		**h =12**	
FGLS	25%	0.218**	(0.05)	0.338**	(0.04)	0.042	(0.16)	0.051	(0.22)
	50%	0.365**	(0.04)	0.507**	(0.03)	0.176	(0.13)	0.220	(0.16)
	75%	0.719*	(0.08)	0.795*	(0.06)	0.417	(0.15)	0.625*	(0.10)
Restricted	25%	0.244**	(0.03)	0.279**	(0.02)	0.123	(0.13)	0.007	(0.32)
FGLS	50%	0.349***	(0.01)	0.382***	(0.01)	0.275**	(0.04)	0.162	(0.13)
	75%	0.091	(0.23)	0.157	(0.15)	0.095	(0.24)	0.078	(0.29)
Panel C : Monthly data									
		h =1		**h =3**		**h =6**		**h =12**	
FGLS	25%	-0.690	(0.20)	-2.417	(0.58)	-3.166	(0.84)	-3.148	(0.62)
	50%	-0.505	(0.17)	-1.834	(0.49)	-4.115	(0.83)	-3.385	(0.45)
	75%	1.022	(0.23)	1.707	(0.18)	-4.458	(0.71)	-2.329	(0.69)
Restricted	25%	0.880*	(0.08)	-2.234	(0.87)	-1.038	(0.80)	0.379	(0.18)
FGLS	50%	1.595**	(0.03)	-1.926	(0.84)	-1.198	(0.76)	1.444	(0.04)
	75%	0.632	(0.21)	0.536	(0.25)	0.297	(0.34)	0.280	(0.29)

Notes: This table reports the R_{OS}^2 statistics for the competitor models, FGLS and restricted FGLS, compared with the historical average benchmark model. The in-sample proportion choices are in the second column. The Clark and West (2007) adjusted MSFE p-values are reported in parentheses, which test the null hypothesis $H_0: \ MSFE_0 \leq MSFE_1$ against $MSFE_0 > MSFE_1$, corresponding to $H_0: \ R_{OS}^2 \leq 0$ against $R_{OS}^2 > 0$. *h* refers to the forecasting horizon. *, **, *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Utility gains from using the FGLS instead of historical average

Panel A : Daily data

	$\gamma = 3$				$\gamma = 6$				Average
	h =1	**h =2**	**h =5**	**h =10**	**h =1**	**h =2**	**h =5**	**h =10**	
25%	2.806	2.752	3.184	3.258	1.418	1.294	1.324	1.436	2.184
50%	5.136	5.041	5.196	4.998	2.566	2.519	2.596	2.497	3.819
75%	4.866	4.714	4.812	4.802	2.431	2.356	2.404	2.400	3.598

Panel B : Weekly data

	$\gamma = 3$				$\gamma = 6$				Average
	h =1	**h =2**	**h =4**	**h =12**	**h =1**	**h =2**	**h =4**	**h =12**	
25%	3.640	4.547	1.881	0.898	1.774	2.043	0.713	0.483	1.997
50%	5.784	6.082	5.149	3.531	2.888	3.037	2.524	1.730	3.841
75%	6.120	6.761	6.202	4.851	3.056	3.377	3.098	2.423	4.486

Panel C : Monthly data

	$\gamma = 3$				$\gamma = 6$				Average
	h =1	**h =3**	**h =6**	**h =12**	**h =1**	**h =3**	**h =6**	**h =12**	
25%	3.824	0.293	1.165	0.406	2.767	-0.410	-0.594	-0.028	0.928
50%	6.607	2.226	2.925	2.728	3.735	-0.153	-0.871	1.077	2.284
75%	8.845	8.044	8.298	0.245	5.291	4.422	1.880	-0.425	4.575

Notes: This table reports the average utility gains of using the FGLS model based on the oil price instead of the historical average model to forecast the stock returns. The utility gain, in an annualised percentage, is the management fee the mean-variance investors are willing to pay for access to the forecasting model. The in-sample proportion choices are in the second column. γ refers to the risk-aversion, and h refers to the forecasting horizons.

Table 7: Restricted FGLS versus historical average at the sector level

		h=1		h=2		h=5		h=10	
Energy	25%	-0.045	(0.80)	-0.033	(0.74)	-0.042	(0.72)	-0.048	(0.74)
	50%	-0.047	(0.73)	-0.032	(0.70)	-0.044	(0.68)	-0.057	(0.74)
	75%	-0.032	(0.65)	-0.011	(0.51)	-0.029	(0.61)	-0.054	(0.740
Material	25%	-0.036	(0.85)	-0.043	(0.87)	-0.051	(0.86)	-0.053	(0.87)
	50%	-0.042	(0.83)	-0.047	(0.85)	-0.059	(0.85)	-0.061	(0.85)
	75%	-0.051	(0.81)	-0.055	(0.80)	-0.071	(0.82)	-0.074	(0.83)
Industrial	25%	0.024	(0.17)	0.016	(0.22)	0.041	(0.11)	0.026	(0.17)
	50%	0.054*	(0.07)	0.047*	(0.08)	0.058*	(0.06)	0.044*	(0.10)
	75%	0.019	(0.27)	0.019	(0.27)	0.025	(0.24)	0.020	(0.28)
Consumer	25%	0.031	(0.30)	0.024	(0.33)	0.033	(0.25)	0.026	(0.29)
Staples	50%	0.027**	(0.02)	0.020*	(0.06)	0.031***	(0.01)	0.023**	(0.05)
	75%	0.010	(0.21)	0.007	(0.28)	0.005	(0.34)	0.004	(0.36)
Healthcare	25%	-0.064	(0.75)	-0.071	(0.80)	-0.050	(0.72)	-0.053	(0.75)
	50%	-0.022	(0.45)	-0.032	(0.53)	-0.022	(0.47)	-0.045	(0.65)
	75%	0.014	(0.30)	0.000	(0.36)	0.000	(0.37)	-0.006	(0.40)
Consumer	25%	0.052**	(0.04)	0.039*	(0.06)	0.068**	(0.03)	0.048*	(0.07)
discretion	50%	0.060**	(0.05)	0.056*	(0.06)	0.064**	(0.04)	0.050*	(0.08)
	75%	0.003	(0.36)	0.001	(0.38)	0.005	(0.35)	-0.003	(0.42)

Telecom	25%	0.014	(0.17)	0.014	(0.16)	0.040**	(0.03)	0.032**	(0.05)
	50%	0.055***	(0.00)	0.056***	(0.00)	0.057***	(0.00)	0.052***	(0.01)
	75%	0.013	(0.18)	0.015	(0.15)	0.010	(0.23)	0.006	(0.33)
Utility	25%	-0.142	(0.43)	-0.133	(0.47)	-0.114	(0.44)	-0.097	(0.41)
	50%	-0.155	(0.44)	-0.147	(0.49)	-0.125	(0.45)	-0.111	(0.43)
	75%	-0.153	(0.70)	-0.135	(0.67)	-0.131	(0.68)	-0.125	(0.67)
Financial	25%	0.025*	(0.10)	0.017	(0.16)	0.019	(0.16)	0.010	(0.24)
	50%	0.041**	(0.05)	0.035*	(0.06)	0.030*	(0.09)	0.026	(0.12)
	75%	0.054**	(0.03)	0.050**	(0.04)	0.048**	(0.04)	0.042*	(0.07)
Technology	25%	0.038*	(0.10)	0.037	(0.11)	0.063**	(0.04)	0.059**	(0.04)
	50%	0.086**	(0.03)	0.083**	(0.04)	0.115***	(0.01)	0.098**	(0.02)
	75%	0.005	(0.35)	0.001	(0.37)	0.010	(0.32)	0.003	(0.36)

Notes: This table reports the R^2_{OS} for the competitor model restricted FGLS compared with the benchmark model historical average for 10 US sectors categorised by GICS. The in-sample proportion choices are in the second column. The Clark and West (2007) adjusted MSFE p-values are reported in parentheses, which test the null hypothesis $H_0: \; MSFE_0 \leq MSFE_1$ against $MSFE_0 > MSFE_1$, corresponding to $H_0: \; R^2_{OS} \leq 0$ against $R^2_{OS} > 0$. *h* refers to the forecasting horizon. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.

Table 8: Utility gains from using the restricted FGLS instead of historical average at the sector level

	$\gamma = 3$				$\gamma = 6$				**Average**
	h=1	**h=2**	**h=5**	**h=10**	**h=1**	**h=2**	**h=5**	**h=10**	
Panel A: in-sample size 25%									
Energy	-3.104	-2.307	-3.029	-2.572	-1.908	-1.605	-1.964	-2.182	-2.334
Material	-2.599	-2.974	-3.356	-3.443	-2.062	-2.236	-2.830	-2.859	-2.795

Industrial	2.505	2.666	3.602	3.186	0.807	0.633	1.040	0.948	1.923
Consumer Staples	1.303	0.907	1.220	1.558	-0.365	-0.437	-0.380	-0.444	0.420
Healthcare	0.189	0.025	0.409	0.608	-0.167	-0.298	0.049	0.098	0.114
Consumer Discretion	3.723	3.277	3.774	3.317	1.658	1.427	1.562	1.387	2.516
Telecom	-0.438	-0.395	1.467	1.095	-0.212	-0.251	0.695	0.555	0.315
Utility	-2.003	-2.066	-2.212	-2.487	-4.078	-4.093	-4.107	-3.893	-3.117
Financial	3.400	2.930	3.106	2.634	1.741	1.543	1.201	1.392	2.243
Technology	1.603	1.413	2.589	2.555	0.799	0.709	1.293	1.284	1.531
Panel B: in-sample size 50%									
Energy	-2.869	-2.135	-2.867	-3.128	-1.783	-1.109	-1.813	-2.350	-2.257
Material	-3.340	-3.827	-4.321	-4.345	-2.599	-2.838	-3.530	-3.607	-3.551
Industrial	5.091	3.124	5.314	3.654	2.544	2.363	2.656	2.453	3.400
Consumer Staples	0.970	0.794	1.119	0.956	0.484	0.396	0.558	0.477	0.719
Healthcare	2.196	1.820	1.907	1.747	1.242	1.039	1.121	0.994	1.508
Consumer Discretion	4.567	4.360	4.751	4.201	2.282	2.178	2.374	2.099	3.352
Telecom	2.235	2.363	2.506	2.309	1.116	1.180	1.251	1.153	1.764
Utility	-2.708	-2.537	-2.793	-2.629	-4.615	-4.613	-4.750	-4.762	-3.676
Financial	5.622	5.289	4.733	4.662	2.810	2.643	2.365	2.330	3.807
Technology	4.061	3.852	5.041	4.490	2.029	1.924	2.519	2.243	3.270
Panel C: in-sample size 75%									
Energy	-2.724	-0.823	-2.668	-4.301	-1.773	-0.492	-1.760	-2.976	-2.190
Material	-6.044	-6.688	-7.637	-7.880	-4.877	-5.191	-6.555	-6.807	-6.460
Industrial	5.893	3.182	6.248	3.795	2.945	2.920	3.122	3.143	3.906
Consumer Staples	0.526	0.490	0.495	0.517	0.261	0.244	0.246	0.257	0.380
Healthcare	6.723	6.003	5.794	6.143	3.701	3.318	3.233	3.362	4.785
Consumer Discretion	3.962	3.807	4.023	3.775	1.979	1.902	2.010	1.886	2.918
Telecom	0.183	0.361	0.291	0.103	0.090	0.179	0.144	0.050	0.175
Utility	-4.725	-4.343	-4.722	-4.814	-7.091	-6.765	-6.689	-6.755	-5.738
Financial	12.627	12.261	11.761	11.360	6.311	6.128	5.878	5.678	9.001
Technology	2.588	2.364	2.876	2.605	1.293	1.180	1.436	1.301	1.955

Notes: This table reports the average utility gains of using the restricted FGLS model based on the oil price instead of the historical average model to forecast the stock returns for 10 US sectors categorised by GICS. The utility gain, in annualised percentage, is the management fee the mean-variance investors are willing to pay for access to the forecasting model. The in-sample proportion choices are in the second column. γ refers to the risk-aversion of investors and h refers to the forecasting horizons.

Table 9: Determinants of predictability at sector level

	BM		DY		MV		PE		TV	
Panel A: in-sample size 25%										
Energy	0.085***	(0.00)	0.208***	(0.00)	-2.621***	(0.00)	-0.054***	(0.00)	0.0021***	(0.00)
Material	0.200***	(0.00)	0.257***	(0.00)	-0.451***	(0.00)	-0.084***	(0.00)	-0.0003	(0.63)
Industrial	0.963***	(0.00)	0.365***	(0.00)	-0.820***	(0.00)	-1.010***	(0.00)	0.0040***	(0.00)
Consumer Staples	0.198***	(0.00)	0.225***	(0.00)	-0.024***	(0.00)	-0.117***	(0.00)	0.0006***	(0.00)
Healthcare	0.202***	(0.00)	0.784***	(0.00)	-0.661***	(0.00)	-0.515***	(0.00)	0.0005	(0.53)
Consumer Discretion	0.799***	(0.00)	0.287***	(0.00)	-1.761***	(0.00)	-0.364***	(0.00)	0.0086***	(0.00)
Telecom	0.002**	(0.02)	0.005	(0.37)	-0.050***	(0.00)	-0.002	(0.28)	0.0000	(0.83)
Utility	0.000	(0.88)	-0.014**	(0.03)	0.058***	(0.00)	0.001	(0.82)	0.0000	(0.75)
Financial	0.219***	(0.00)	0.128***	(0.00)	-2.070***	(0.00)	-0.030***	(0.00)	0.0000	(0.91)
Technology	0.174***	(0.00)	-0.013**	(0.04)	-4.187***	(0.00)	-0.105***	(0.00)	0.0051***	(0.00)
Panel B: in-sample size 50%										
	BM		DY		MV		PE		TV	
Energy	0.133***	(0.00)	0.223***	(0.00)	-3.005***	(0.00)	-0.046***	(0.00)	0.0016**	(0.04)
Material	0.388***	(0.00)	0.185***	(0.00)	-0.777***	(0.00)	-0.106***	(0.00)	-0.0001	(0.87)
Industrial	1.016*	(0.0	0.690	(0.0	-	(0.0	-	(0.0	0.0013*	(0.0

	**	0)	***	0)	0.625***	0)	1.212***	0)	*	7)
Consumer Staples	0.200***	(0.00)	0.228***	(0.00)	-0.024***	(0.00)	-0.126***	(0.00)	0.0004**	(0.06)
Healthcare	0.669***	(0.00)	1.111***	(0.00)	-0.966***	(0.00)	-0.725***	(0.00)	0.0009	(0.42)
Consumer Discretion	1.104***	(0.00)	0.295***	(0.00)	-1.641***	(0.00)	-0.352***	(0.00)	0.0085***	(0.00)
Telecom	0.001	(0.37)	0.002	(0.71)	-0.022***	(0.00)	0.000	(0.92)	0.0000	(0.97)
Utility	-0.001	(0.79)	-0.014**	(0.05)	0.061***	(0.00)	0.003	(0.41)	-0.0001	(0.30)
Financial	0.292***	(0.00)	0.128***	(0.00)	-2.152***	(0.00)	-0.010	(0.34)	0.0001	(0.66)
Technology	0.267***	(0.00)	-0.003	(0.60)	-4.135***	(0.00)	-0.156***	(0.00)	-0.0043***	(0.00)

Table 9 continued

	BM		DY		MV		PE		TV	
	Panel C: in-sample size 75%									
Energy	0.032**	(0.02)	0.392***	(0.00)	-3.280***	(0.00)	-0.033	(0.30)	0.0068***	(0.00)
Material	0.285***	(0.00)	0.091***	(0.00)	-1.627***	(0.00)	-0.025	(0.20)	-0.0013	(0.12)
Industrial	0.365***	(0.00)	0.052*	(0.10)	-2.818***	(0.00)	-0.302***	(0.00)	-0.0010	(0.32)
Consumer Staples	0.143***	(0.00)	0.038***	(0.00)	-0.625***	(0.00)	-0.048***	(0.00)	0.0004*	(0.08)
Healthcare	1.183***	(0.00)	0.429***	(0.00)	-1.946***	(0.00)	-0.393***	(0.00)	-0.0036***	(0.00)
Consumer Discretion	0.670***	(0.00)	0.045	(0.23)	-2.124***	(0.00)	-0.261***	(0.00)	0.0006	(0.39)
Telecom	-0.028***	(0.00)	-0.011**	(0.09)	0.022***	(0.00)	0.002	(0.66)	0.0001	(0.49)
Utility	0.000	(0.99)	-0.006	(0.29)	0.027***	(0.00)	0.001	(0.88)	-0.0001	(0.30)
Financial	0.279***	(0.00)	0.072***	(0.00)	-2.115***	(0.00)	-0.008	(0.48)	-0.0004	(0.17)
Technology	0.139***	(0.00)	-0.023***	(0.01)	-4.251***	(0.00)	-0.043	(0.18)	0.0073***	(0.00)

Notes: This table reports results on the determinants of predictability by sector. Time series data on predictability is computed using the difference of forecasted error from historical average and FGLS models $d_{t+1} = (r_{t+1} - \hat{r}_{0,t+1})^2 - (r_{t+1} - \hat{r}_{1,t+1})^2$; $\hat{r}_{0,t+1}$ is the forecasted equity premium of the historical average model and $\hat{r}_{1,t+1}$ is from the FGLS model. The time series data on predictability are generated from out-of-sample forecasting with in-sample sizes of 25%, 50% and 75%. The determinants for equity premium predictability are sectoral characteristics including book-to-market ratio (BM), dividend yield (DY), market capitalisation (MV), trading volume (TV), and price-earnings ratio (PE). All independent variables are taken in natural logarithmic form. A unit root test reveals that except for the predictability variable, all other variables are non-stationary; therefore, we take the first difference of these variables and fit a GARCH (1, 1) model. The estimated coefficients together with their p-values (in parentheses) are reported. *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.