Can Oil Prices Help Predict US Stock Market Returns: An Evidence Using a DMA Approach

Naser, Hanan and Alaali, Fatema

Department of Economics, University of Sheffield, UK, Department of Economics, University of Sheffield, UK

19 January 2015

Can Oil Prices Help Predict US Stock Market Returns: An Evidence Using a DMA Approach

Hanan Naser*, Fatema Alaali

Department of Economics, University of Sheffield, UK

June 26, 2015

Abstract

Crude oil price behaviour has fluctuated wildly since 1973 which has a major impact on key macroeconomic variables. Although the relationship between stock market returns and oil price changes has been scrutinized excessively in the literature, the possibility of predicting future stock market returns using oil prices has attracted less attention. This paper investigates the ability of oil prices to predict S&P 500 price index returns with the use of other macroeconomic and financial variables. Including all the potential variables in a forecasting model may result in an over-fitted model. So instead, dynamic model averaging and dynamic model selection are applied to utilize their ability of allowing the best forecasting model to change over time while parameters are also allowed to change. The empirical evidence shows that applying the DMA/DMS approach leads to significant improvements in forecasting performance in comparison to other forecasting methodologies and the performance of these models are better when oil prices are included within predictors.

Keywords: Bayesian methods, Econometric models, Macroeconomic forecasting, Kalman filter, Model selection, Dynamic model averaging, Stock returns predictability, Oil prices

JEL classification: C11, C53, G17, Q43

*Corresponding author E-mail Address: ecp10hsn@sheffield.ac.uk

1 Introduction

It is not surprising to see a surge in oil market research. According to the International Energy Agency (IEA), oil covers 31.4% of the total primary energy supply and it is projected to supply 28.5% of the world's total energy needs (IEA, 2014). Moreover, oil prices have witnessed wide swings in times of shortage or oversupply as the crude oil price cycle may extend over several years due to a change in demand or geopolitical events. This fluctuation in oil prices seems to have a great effect on the economy in general and in particular on investor's portfolios as the oil and stock markets are cross-hedged (Arouri et al., 2011). Oil prices instability can affect stock returns through different channels. For instance, an increase in the price of oil, which, in the absence of the effects of entire substitution between the components of production, increases the cost of operating a business and hence reduces cash flow. As it is known, stock prices are discounted values of expected cash flows, therefore, a reduction in the cash flow causes a similar change in stock prices (Huang et al., 1996). Another indirect channel that transmits the impact of oil price changes to stock returns is the discount rate as it consists of a combination of the expected inflation rate and expected real interest rate, both of which may be affected by the price of oil. Like all other commodities, rising oil prices are often indicative of inflationary pressures. Therefore, an increase in the expected inflation rate will cause the same change in discount rate, thus, a reduction in stock returns (Basher et al., 2012).

Accordingly, an excessive number of studies scrutinized the relationship between oil price changes and stock market returns arguing that if the real economy is affected by oil price shocks through consumer and firm behaviors, then stock market returns would be affected (see, e.g., Basher et al., 2012, Chang and Yu, 2013, Chen, 2010, Cunado and de Gracia, 2014, Driesprong et al., 2008, Jones and Kaul, 1996, Kilian and Park, 2009, Miller and Ratti, 2009, Mollick and Assefa, 2013, Park and Ratti, 2008, Sadorsky, 1999). However, the main concern of investors and market participants is the impact of oil price shocks on their portfolios. Therefore, Jones and Kaul (1996) suggest employing the changes in oil price in forecasting stock market returns. The studies mentioned above relied on in-sample investigation, however, given that out-of-sample exercises provide a measure of protection against data mining, as those observations are not used in estimation, it is interesting to examine the performance of forecasting technique on the basis of out-of-sample analysis. Hence, Narayan and Gupta (2015) use over 150 years of monthly data to predict US stock returns using West Texas Intermediate (WTI) crude oil price. Their out-of-sample forecasting results show that oil price is a persistent predictor variable. To enrich this field of research, the first characteristic of this paper is to scrutinize the ability of oil price changes in forecasting US stock returns in line with other macroeconomic and financial variables.

In a recent survey by Rapach and Zhou (2013), a comprehensive investigation on the forecasting performance of stock returns in literature, three key points are made. First, as there is a strong relationship between stock returns predictability and business cycle fluctuations, the performance of the forecasts differ in economic expansions from that in recessions. This implies that different predictors may be required to explain stock market returns at different points in time. For example, Pesaran and Timmermann (1995) investigate the robustness of predicting US stock returns using forecasting models that allow the set of predictors to change over time and show its importance in improving forecast performance. They declare that when the same forecasting model is used over the whole period, the model will be criticized for ignoring the problem of model uncertainty, which is a strong assumption. Subsequently, Avramov (2002) and Cremers (2002) give rise to model uncertainty by using Bayesian model averaging (BMA) and Bayesian model selection (BMS), respectively. Second, parameter instability may cause the best model to change over time as the relationships between economic variables may not be constant but change over time (see. e.g., Groen et al., 2013, Sarno and Valente, 2009, Pesaran and Timmermann, 2000, 2002). As the change in oil prices is our main interest, the literature has detected its time-varying relationship with stock market returns (Kilian and Park, 2009, Miller and Ratti, 2009). Third, recent strategies that deliver statistically and economically significant out-of-sample gains, including strategies based on economically motivated model restrictions (e.g., Campbell and Thompson, 2008), forecast combination (e.g., Rapach et al., 2010), diffusion indices (e.g., Ludvigson and Ng, 2007), and regime shifts (e.g., Henkel et al., 2011) clearly improve forecasting performance by accounting for model uncertainty and parameter instability.

Therefore, to address all of the above issues, Koop and Korobilis (2011, 2012) introduced dynamic model averaging (DMA) strategy that allows for the best forecasting model to change over time while parameters, at the same time, are also allowed to change. This approach allows using a large number of predictors (m) to generate $K = 2^m$ models, which are characterized by having different subsets of explanatory variables. The DMA terminology is to do model averaging for the different models that hold at each point of time. Alternatively, a single model can be selected at each point of time on the basis of its performance in order to do dynamic model selection (DMS). Although, the DMS approach has been applied by Liu et al.

(2015) who employ oil market variables to forecast S&P 500 stock returns, this paper applies both, DMA and DMS approaches, which makes this study, to the best of our knowledge, the first to consider both parameters and model uncertainty in forecasting stock market returns.

One final contribution of this study is that recursive estimations have been utilized over the out-of-sample period to compute multi-steps-ahead estimations for the returns of US stock market. Our out-of-sample results indicates that applying the DMA/DMS approach leads to significant improvements in forecasting performance in comparison to other forecasting approaches. Given that both, DMA and DMS methodologies shrink models in different ways, our results prove that the DMS add more gains to the forecasting performance in comparison to the DMA. Furthermore, investigating different specifications of DMA/DMS models shows that forecasting US stock market returns has model uncertainty rather than parameter uncertainty. Finally, the performance of the DMA/DMS models are better when including oil prices within predictors.

The rest of the paper is organized as follows. Section 2 explains the econometric methodology followed to predict US stock returns. The variables used and their sources are summarized in section 3. Empirical work and the discussion of the results obtained are presented in section 4. Section 5 reports the main conclusion of this study.

2 Econometric Methodology

To introduce the importance of incorporating variables that describe stock market returns behavior, a simple regression model is applied based on carefully selected variables:

$$y_t = \alpha + \sum_{k=1}^{p} \phi_k y_{t-k} + \sum_{k=0}^{p} \phi_k X_{t-k} + \varepsilon_t \tag{1}$$

where y_t is the stock market returns and X represents a matrix that comprises explanatory variables, which describe stock market returns behavior.

Alternatively, this paper considers a number of different models in forecasting stock market returns including: the time-varying parameter (TVP) models, the dynamic model averaging (DMA), the dynamic model selection (DMS), and the Bayesian model averaging (BMA).

In general, the key advantage of using time-varying parameter (TVP) models is that these models account for parameter uncertainty. They utilize state space methods such as the Kalman filter, which is used by Koop and Korobilis (2011,

2012) in empirical macroeconomic forecasting. However, such models fail to account for model uncertainty, and thus do not allow the set of predictors to change over time (Koop and Korobilis, 2012). Accordingly, if the number of potential predictors is large, then the TVP models tend to over fit in-sample, and hence have a poor out-of-sample forecasting performance. To overcome these limitations in the TVP models, DMA models provide a superior alternative.

To explain the way that dynamic model averaging (DMA) works, we will focus on the two main advantages of this approach named parameter and model instability. Starting with parameter instability, the time varying parameter (TVP) models which are estimated using state space methods as discussed above can be written as the following:

$$\begin{aligned} y_t &= Z_t \theta_t + \varepsilon_t \\ \theta_{t+1} &= \theta_t + \eta_t \end{aligned} \tag{2}$$

where in our case, y_t is the stock price returns and Z_t is an $1 \times m$ vector of an intercept, all explanatory variables, and lagged values of y_t.[1] θ_t is an $m \times 1$ vector of coefficients, $\varepsilon_t \sim i.i.d.\ N(0, H_t)$ and $\eta_t \sim i.i.d.\ N(0, Q_t)$. Such time-varying models can be estimated using standard methods involving a Kalman filter and smoother (see Koop, 2003, Cogley and Sargent, 2005, Justiniano and Primiceri, 2008). However, the model in equation (2) assumes that the set of predictors included in Z_t remains constant at all points of time. This might be a strong assumption as explained in the introduction. Also, despite the computational complexity, if the number of explanatory variables in Z_t is large, such models often over fit in sample and cause poor forecasts. Empirical evidence provided by Koop and Korobilis (2012, 2011) also shows that maintaining the same forecasting model over time performs poorly due to over-parameterization problems. As a result, we adopt their strategy and allow for K models which utilize different sets of predictors to be applicable at different points of time as shown below:

$$\begin{aligned} y_t &= Z_t^k \theta_t^k + \varepsilon_t^k \\ \theta_{t+1}^k &= \theta_t^k + \eta_t^k \end{aligned} \tag{3}$$

where $Z_t^k \subseteq Z_t$, for $k = 1, 2..., K$, $\varepsilon_t \sim i.i.d.\ N(0, H_t)$ and $\eta_t \sim i.i.d.\ N(0, Q_t)$. The state-space model presented in equation (3) allows for a different best performing model to hold at each point of time, to do model averaging and to select the best performing model. To be precise, forecasting at time t can be proceeded by exploiting information through time $t-1$ and calculating the posterior model probability

[1] Explanatory variables are all transformed to be stationary. For more details, see Section (3).

for individual model k at each point in time in the form of $Pr_t(k \mid y^{t-1})$. This probability is then used to forecast using Dynamic Model Averaging (DMA), Bayesian Model Averaging (BMA) or Dynamic Model Selection (DMS). For DMS, the model with highest probability is selected, while alternatively, these probabilities are used as weights when constructing average forecasts using all models K at each time point t. Basically, for both Dynamic Model Averaging (DMA) and Bayesian Model Averaging (BMA), it is possible to calculate the predictive density, $p(y\hat{}\,|_y)$, by averaging results overall models with weights given by posterior model probability as shown below:

$$p(\hat{y}\,|_y) = \Sigma_{k=1}^{K} p(\hat{y}\,|_{y,k}) p(k \mid y) \tag{4}$$

This method is called Bayesian Model Averaging (BMA) as noted by Raftery et al. (2010). For m number of predictors, we can construct $K = 2^m$ models. After constructing static models with parameter uncertainty, an average forecast is calculated for all the 2^m models, where all observations carry the same weight throughout all the estimation period. In the same way, as there is no forgetting, DMS is the static Bayesian Model Selection (BMS). Alternatively, the DMA modifies the BMA by introducing two important new forgetting factors named parameter forgetting, λ, and model forgetting, α.

The fundamental shortcoming of model (3) is how to compute the evolution of models over time. More concretely, a random variable $S_t \in \{1, 2, ..K\}$ shows the model applied at time t. The random variable S_t is assumed to form a Markov chain with transition probability matrix $P = (p_{ij})'_{ij \notin \Lambda}$. The transition probability $p_{ij} = P(S_t = j|S_{t-1} = i)$ is the probability that the forecasting model at time $t-1$ is i and will switch to model j at time t. Such Markov switching models have been introduced to economics by Hamilton (1989) and have been widely used in economics and finance since then. However, based on this framework the size of transition probability matrix will become computationally infeasible even if the number of models is small. Koop and Korobilis (2012, 2011) get around the curse of dimensionality by using an approximation method suggested by Raftery et al. (2010).

Before explaining the main ideas of the algorithm developed by Raftery et al. (2010), it is worth noting that Bayesian estimates of a state-space model involves Markov Chain Monte Carlo (MCMC) methods which take draws of the states conditional on other parameters such as that of H_t and Q_t (i.e., $\theta_t^k|H_t, Q_t$). Then draw the other parameters conditional on the states.[2] With the large number of models

[2]For a complete description of Bayesian estimation of state-space model see Koop (2003) and Kim and Nelson (1999).

estimated in our application the computation of MCMC will be impossible. The key aspect of Raftery et al. (2010) algorithm is to avoid MCMC by obtaining a plug-in estimate of H_t and assuming $Q_t = (1 - \lambda^{-1})\Sigma_{t-1}$ where $0 < \lambda \leq 1$ and $\Sigma_t = (\theta_t - \widehat{\theta}_t)(\theta_t - \widehat{\theta}_t)'$. Note that $\widehat{\theta}_t$ is the Kalman filter estimate of θ_t and λ is known as a forgetting factor in the sense that observations of j periods in the past have a weight of λ^j. Values of λ close to one suggest high parameter persistence. More concretely, $\lambda = 1$ implies that parameters remain constant. Alternatively, as $\lambda \to 0$, only the last observation is used for forecasting.[3]

The second approximation of Raftery et al. (2010) algorithm concerns the efficient computation of posterior model probabilities. Let $\pi_{t|t-1,k}$ denotes the probability that model k is applied at time t using information up to time $t-1$. $\pi_{t|t-1,k}$ can be used either to do model averaging or select the best forecast performing model. Hence, DMA uses $\pi_{t|t-1,k}$ to weight K different models and DMS selects the model with the highest $\pi_{t|t-1,k}$. When using Markov switching process to describe the evolution of K models with transition probability P and the predictive density of model k given by $p_k(y_{t-1}|y_{t-2}, y_{t-3}...y_1)$ then

$$\pi_{t|t-1,k} = \sum_{i=1}^{K} \pi_{t-1|t-1,k} p_{ij} \tag{5}$$

where

$$\pi_{t-1|t-1,k} = \frac{\pi_{t-1|t-2,k} p_k(y_{t-1}|y_{t-2}, y_{t-3}...y_1)}{\sum_{l=1}^{K} \pi_{t-1|t-2,k} p_l(y_{t-1}|y_{t-2}, y_{t-3}...y_1)} \tag{6}$$

However, it is noted above that such a strategy is computationally impossible because P is too large even for cases where K is moderately large. Raftery et al. (2010) circumvent this problem by replacing (5) by

$$\pi_{t|t-1,k} = \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l}} \tag{7}$$

where $0 < \alpha \leq 1$ is another forgetting factor with interpretation similar to λ but in terms of model rather than parameter evolution. The interpretation of α becomes apparent if equation (7) is written as:

$$\pi_{t|t-1,k} \propto \prod_{i=1}^{t-1} [p_k(y_{t-i}|y_{t-i-1}...1)]^{\alpha^i} \tag{8}$$

[3]As λ decreases, a greater and greater degree of coefficients change is allowed. As $\lambda \to 0$, only most recent observations is used for forecasting.

It can be seen that values of α close to one imply that $\pi_{t|t-1,k}$ will be larger and the DMS will select model k at time t if it forecasted well in the recent past.[4] In other words, at each point of time there will be different values of $\pi_{t|t-1,k}$ corresponding to each model. The DMS proceeds by selecting the model with highest value of $\pi_{t|t-1,k}$ rather than averaging all models such as that of DMA.

In the empirical work of this study, Raftery et al. (2010) is followed by using Kalman filtering and smoothing methods for fast real time forecasting. In general, equation (3) is estimated by obtaining a plug in estimate of H_t, then replacing it by $\hat{H}_t$ and assuming $Q_t = (1 - \lambda^{-1})\Sigma_{t-1}$, where $0 < \lambda \leq 1$ and Σ_{t-1} is the covariance matrix of the estimation error in Kalman filter (i.e the estimation error is $(\theta_t - \hat{\theta}_t)$ where θ_t is the filter estimate)[5]. As monthly data are used in this study, it is expected to have relative frequent change in both parameters and models involved in the forecasting exercise. Given that there is no conscious in literature with regards to the best values of λ and α, Raftery et al. (2010) propose to set these values at 0.99. In addition, Koop and Korobilis (2011, 2012) implemented a grid search instead, and present robustness checks for various combinations of λ and α with lowest and highest values of 0.93 and 1.0, respectively. They find that in the impact of the size of the forgetting factors to be relatively small in forecasting the US inflation rate, and thus recommend to set the values at 0.99. Hence, following Koop and Korobilis (2012), the values of the forgetting factors are fixed to $\lambda = 0.99$ and $\alpha = 0.99$.[6] Since BMA is a special case of DMA model with $\lambda = \alpha = 1$, and DMA and DMS; with $\lambda = 1$ and $\alpha = 0.99$, all relax parameters variation, so, getting better forecasting performance for these models provide information about the importance of accounting for model uncertainty rather than parameters instability in forecasting US stock returns. On top of that, the DMA and DMS models provide information on the importance of each explanatory variable at each point of time. Alternatively, models that allow parameters to change but the forecasting model to remain constant are used to examine whether this can help improve the performance of forecasting stock market returns (i.e. TVP-AR (1) and TVP-SR with $\lambda = 0.99$). In order to compare the performance of the main models, other alternative benchmark models

[4]For instance, if monthly data are used with $\alpha = 0.99$ then the forecasting model used three years ago receives around 70% as much weight as the forecasting model used last period. If $\alpha = 0.95$ then forecast performance three years ago receives only 16% weight.

[5]Raftery et al. (2010) explain in details the use of this approximation

[6]Note that, Grassi and de Magistris (2015) suggest that one could choose values for λ and α on the basis of the forecast performance, however; this would not only bias the findings in favor of DMA, it is also not a valid procedure for the out-of-sample forecasting (Koop and Korobilis, 2012). Hence, following Koop and Korobilis (2012), the values for the forgetting factors are simply selected.

have been constructed.[7]

3 Data

In this paper, US stock market returns are predicted by employing monthly returns on the S&P 500 price index over the period January 1959 to December 2013. The data for the S&P 500 price index and other macro-variables are acquired from the International Financial Statistics (IFS) issued by the International Monetary Fund (IMF), however, West Texas Intermediate (WTI) spot oil price data are attained from FRED II (Federal Reserve Economic Data). Following Rapach et al. (2005) and Chen (2009), the variables used to predict stock returns are:

- industrial production growth measured by first difference of the log values of industrial production index;
- term spread measured as the difference between the 10-year government bond yield and the 3-month constant maturity Treasury bill rate;
- inflation rate proxied by the first difference of the log values of the consumer price index;
- narrow money growth (M1) measured by the first difference of the log values of M1;
- broad money growth (M2) measured by the first difference of the log values of M2;
- interest rate measured as the change in federal funds rate;
- change in unemployment rate;
- real oil price returns measured as the change in the log values of West Texas Intermediate (WTI) spot oil price.

Moreover, the following fundamental variables from Welch and Goyal (2008) and Campbell and Thompson (2008), that are downloaded from Shiller (2015) site, are included.

[7] Since Rapach et al. (2005) have used a univariate AR process to evaluate the performance of stock returns predictability without any information from other variables, this paper employs three classical models (i.e. with no TVP) including the random walk, single autoregressive, and multivariate simple regression model, to evaluate the forecast performance of the DMA and DMS models as shown in section 4.

- dividend yield is the difference between the log of dividends and the lagged log stock prices, where dividends are measured using a one-year moving sum.

- earning price ratio is the difference between the log of earnings and the log of stock prices, where earnings are measured using a one-year moving sum.

- book to market ratio is the ratio of the Dow Jones Industrial Average book to market values.

Table 1 shows the mean and standard deviation of the variables used. Three unit root tests are applied to all the variables; the Augmented Dickey and Fuller (1979) (ADF), Kwiatkowski et al. (1992) (KPSS) and Phillips and Perron (1988) (PP) tests. The obtained results of all three tests are presented in Table 2.[8] ADF and PP tests results are significant, which translate to rejection of the null hypothesis of unit root whereas the KPSS test results are insignificant which denote that the stationary null hypothesis cannot be rejected. The full sample that covers the period from January, 1959 to December, 2013, is divided it into two sub-samples. Given that the stock returns (SR) is defined as the month-on-month percentage change in the S&P 500 index and the set of predictors are transformed to ensure stationarity, the effective sample starts from February, 1959. The first 300 observations are used for estimation while the out-of-sample forecasts are carried out from February, 1984 to December, 2013, which covers the recent boom and bust periods. The Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) have been used to select the number of lags included in AR(p) and DMA/DMS models. In this set up, the total number of potential predictors in the DMA approach becomes 11, which could lead to generate ten thousands (i.e. 2^{11}) different models, and accordingly create computational difficulties through out the estimation process. Therefore, the problem of dimensionality concerning the number of potential predictors is controlled by implementing the DMA and DMS approach suggested by Koop and Korobilis (2012).

Since Hamilton (2011) has quoted clearly that "...deflating by a particular number such as the CPI introduces a new source of measurement error, which could lead to a deterioration in the forecasting performance. In any case, it is again quite possible that there are differences in the functional form of forecasts based on nominal instead of real prices" (Page 370), using real variables causes measurement errors. Thus, Narayan and Gupta (2015) has been followed in this paper, where stock returns and oil price changes are used in nominal phase instead of real variables.

[8]ADF and PPerron are the Augmented Dickey Fuller and Phillips-Perron tests with the null hypotheses of unit root and KPSS test is based on a null hypothesis of stationary time series.

Given that oil price return is the main variable under consideration, Figure 1 shows a plot of WTI oil price and S&P 500 price index over the period under study. It can be seen that the steady nature of oil price ends in the year 1973 when a significant energy crisis started as a result of the embargo imposed by Arab oil producers that cause a rise in the price of crude oil from \$3 per barrel to \$12. This was followed by a loss in production that was caused by the combined effect of the Iranian revolution and Iran-Iraq war which resulted in a rise in crude oil prices to more than double from \$14 to \$35 in 1981. The price of oil spiked again in summer 1990 due to the uncertainty associated with the Iraqi Invasion of Kuwait followed by Gulf war. The 1973-1974 period was the worst stock market crash, recovery of the countries was very slow. The United states only returned to the same level in 1993. The rise in oil prices came to a rapid end in 1998 as a consequence of higher OPEC production. After this period, both oil prices and the S&P 500 price index have mostly climbed. There has been a very strong positive correlation between the two since the second half of 2008. This may be partly attributed to the portion of the energy sector that accounts for 12.3% of the market capitalization of the S&P 500. Moreover, the S&P 500 energy sector tends to outperform (underperform) the S&P 500 when the price of oil is rising (falling). According to Kilian and Park (2009), the response of stock market returns differs greatly depending on the source of oil price shocks. Rise in oil prices that is caused by a specific oil demand shock will probably be accompanied by a drop in stock prices. In contrast, disruption in crude oil supply has no significant effect. Unexpected global economic expansion that drive oil prices up, has a constant positive impact on stock prices.

4 Empirical Results

In order to generate out-of-sample predictions for stock returns, recursive window has been used as suggested by Rapach et al. (2010) to apply the estimated regression function to the observations that were not used to generate the estimates. Virtually, the sample was first divided into two sub-samples, where the first part covers T months and represents the in-sample observations, while the rest shows the out-of-sample period. Accordingly, the observations in the in-sample period are used to obtain forecast of stock returns on the T + h month. Then the observations for the initial T + h months are used to obtain the return forecasts on the next month an so on. Carrying out the forecasts using the same method through the end of the out-of-sample period, a series of n out-of-sample forecasts of stock returns is generated. In this paper, the initial in-sample period covers the period from February 1959 to January 1984 and the evaluation period starts from February 1984 to December 2013.

The empirical results of using DMA and DMS models are presented into two main subsections. The first subsection compares the forecast performance of the DMA/DMS models to a number of alternative forecasting models, while the second subsection provides evidence of which variables are more important in predicting the stock market returns and whether or not they change over time . The considered forecast horizons are: one ($h = 1$), three ($h = 3$), six ($h = 6$) and twelve ($h = 12$).

4.1 Forecasting performance

To properly analyze the accuracy of forecasts provided by the DMA and DMS models, the mean squared forecast error (MSFE) in percentages is used to compare the performance of all the models. MSFE's are reported over the out-of sample period which starts from February 1984 to December 2013. In terms of alternative forecasting models, the results for the models below are reported:

1. Forecasts using random walk (RW) model;
2. Forecasts using first order autoregressive (AR(1)) model with an intercept;
3. Forecasts using all variables in a simple regression (SR) model with an intercept;
4. Forecasts using first order autoregressive model with time-varying coefficients (i.e., TVP-AR(1)) at $\lambda = 0.99$;
5. Forecasts using all variables in a simple regression model with time varying parameters (i.e., TVP-SR)at $\lambda = 0.99$;
6. Forecasts using DMA and DMS with $\alpha = \lambda = 0.99$;
7. Forecasts using BMA as a special case of DMA (i.e., $\alpha = \lambda = 1$);
8. Forecasts using DMA and DMS with $\alpha = 0.99$ and $\lambda = 1.0$ (coefficients do not vary over time);

In the first model, the random walk (RW) model is applied, which is a non-Bayesian model and needed as a benchmark. Then, for better comparison, two other classical (i.e. non Bayesian) models are utilized including the first order autoregressive (AR(1)) and the simple regression (SR) model. In the former, the stock market returns are regressed on its lags, where the lags have been selected on the basis of AIC without any further information, while the second predicts the stock returns using rich information as discussed in Section (3). Then, to test whether TVP models improve the forecasting performance, TVP- AR and TVP- SR models are applied at $\lambda = 0.99$. The sixth model is the focus of this study, where the dynamic model averaging (DMA) and dynamic model selection (DMS) are performed using the benchmark values of $\alpha = \lambda = 0.99$.[9] This allows not only the parameters to change over time, but also the set of predictors. As discussed in Section (2, values of 0.99 are consistent with fairly stable models with gradual coefficient changing over time, where forecasts are averages across models using the associated probabilities for each individual model for the DMA, and the DMS involves selecting a single model with the highest probability and using this to forecast stock returns. The BMA is obtained by setting both forgetting factors to one, which allows for using conventional linear forecasting models with no time variations in the coefficients. Finally, although the last model is a constant parameter model, it allows for model evolution, which enables us to have a better picture whether forecasts can be improved using TVP, model uncertainty or both.

Table 3 shows the results obtained by applying the above forecasting exercises for four different horizons; h=1, h=3, h=6, and h=12. The key message of these results is that the DMA and DMS perform well in forecasting stock returns, with DMS being the best overall. Despite the fact that both DMS and DMA do the shrinkage using the fixed forgetting values in several methods, the DMS pay more attention to the one best model as it puts zero weights on all other. Hence, shrinking the contribution of all models except one towards zero. This additional shrinkage provides more benefits to the DMS forecasts over DMA, which is in line with the findings of Koop and Korobilis (2012). Furthermore, in times of rapid changes, DMS will tend to switch more quickly than DMA since it can select an entirely new model as opposed to adjusting the weights on all the models. This reveals that the shrinkage provided by the DMS models provides more benefits in forecasting over the DMA. However, there is model rather than parameter uncertainty in this forecasting exercise. This can be concluded by looking at the values of MSFE provided in Table 3, where using TVP models for forecasting have limited the improvement of the forecast performance on average. For instance, although the MSFE of the models that apply time varying parameters including TVP-AR (1) and TVP-SR are a bit better than earlier classic models MSFE, they result in poorer forecasting performance, relative to the DMA and DMS. In addition, since all the models that include the set of

[9]For more information on selecting the values of forgetting factors, see Koop and Korobilis (2012).

predictors outperforms the RW and AR model model, it is clear that predictors matter in forecasting. However, to obtain better forecast performances, one does not only need to include information from the fundamentals, but also allows for model evolution and parameter evolution as carried out by the DMA and DMS. Interestingly though, based on the results provided by the MSFE, it is obvious that most of the improvements in forecast performance found by DMA or DMS are due to model evolution rather than parameter evolution since DMA and DMS do not perform better than the DMA model with λ set equal to 1 (implying that parameters are not allowed to change here). It can therefore be concluded that allowing for model uncertainty and not only parameter uncertainty, improves the forecasting performance of these models. To examine the impact of including the oil prices as a potential predictor in this study, the best forecasting exercise using the DMA and DMS approach is repeated with the exclusion of oil prices variable. Not surprisingly; despite the fact that DMA and DMS are still better than all other alternative models, results of this workout shows that the efficiency is less than the later best model, which denotes that oil prices are good predictors for stock market returns and it can help improving the forecast performance in general. This result is in line with the finding of Driesprong et al. (2008), who illustrate that changes in oil prices can significantly predict stock market returns in both emerging and developed markets.

4.2 Good predictors for stock market returns

Given that the DMA has a large potential benefits over other forecasting approaches, where it allows both the set of predictors (i.e. forecasting model) and the model parameters' to change overtime, many of the models under consideration are parsimonious as they use a small set of predictors. Thus, if the DMA attached a great deal of weight to such models, it can avoid over-fitting problems. Considering the difficulty in analysing the returns on stock prices, this study uses a set of 11 potential explanatory variables (excluding the lag of the dependent variable), which implies that a total number of 8192 possible models are constructed to choose from. Accordingly, the DMA model weights the probability when constructing average forecasts using all K models (i.e. 8192) at each point of time t, while alternatively the DMS select the 'best' forecasting model with highest probability at each point of time. Precisely, if the $Size_{k,t}$ is the number of predictors in model k at time t then

$$E(Size) = \sum_{k=1}^{k} \pi_{t|t-1} Size_{k,t} \tag{9}$$

is the expected number of predictors included in the DMA at time t. Figure 2 provides the expected number of predictors selected by the DMA model for the stock returns for each forecast horizon, $h = 1$, $h = 3$, $h = 6$ and $h = 12$, over the out-of-sample period, February 1984 to December 2013. The results show that for certain periods and forecast horizons, it might be argued that the DMA approach favors parsimonious models. Although the maximum number of indicators is 11, the maximum number of predictors used to predict stock market returns across all horizons is 9. More precisely, Figure 2 shows that at h = 1, the DMA uses 8 predictors to forecast stock market returns and it does not change noticeably over time. At the longer horizons of $h = 3$ and $h = 6$, the shrinkage of DMA is particularly impressive. The DMA has basically includes between 5 to 7 of the 11 predictors listed in Data section among the out-of-sample period, except for year 2008. However, looking at the expected number of predictors in forecasting exercise of 12 steps ahead, it slightly uses more predictors in modeling the system of DMA approach, but almost never more than 9 predictors are included, showing that the DMA approach is strongly favoring parsimonious models.

Although Figure 2 illustrates the expected number of predictors at each forecasting horizon as well as each point of time t, it does not provide any information about the level of importance of each explanatory variable used. Therefore, plotting the posterior inclusion probabilities for each explanatory variable (eleven in our case) is helpful to indicate which variables are most important through out the examined period. With a glance at the results obtained in Figures 3 to 6 on the main predictors as selected by the DMA analysis for the stock returns for every forecast horizon, $h = 1$, $h = 3$, $h = 6$, and $h = 12$, over the out-of-sample period from February, 1984 to December, 2013, it shows clearly that the importance of each indicator varies obviously over time and forecast horizons.

The Federal fund rate (FFR), term spread (TS) and inflation rate (PI) have never exceeded a probability of 0.5 for all horizons, except for $h = 12$, where it comes through strongly for both term spread and inflation rate. Industrial production (IP) shows strong predictive power throughout the first forecast horizon, while both narrow money (M1) and broad money (M2) measures come through strongly for $h = 12$. It is observed that earning price ratio (EP) has a strong predictive power for $h = 3$, while the unemployment rate receives a higher importance for $h = 12$. Book to market ratio (BM) and oil prices (WTI) exhibit very strong information that can predict stock returns for all horizons. Precisely, for $h = 1$, the BM sticks at 1 for most of the period, while the WTI prices show significant role in predicting stock returns for $h = 12$. In sum, all the predictors show strong predictive power at one time or another though at varying magnitudes, while BM and WTI appear to be strong at almost all horizons among the out-of sample period.

5 Conclusions

This paper investigates the ability of the change in oil prices to predict US stock returns with the use of other macroeconomic variables such as industrial production, the term spread, money stocks, interest rate, inflation rate, unemployment rate and financial ratios. Both in-sample estimation and out-of-sample forecasts are implemented. The main finding is that DMA and DMS approaches are more effective in forecasting macroeconomic variables. These approaches allow for the best forecasting model to change over time while parameters, at the same time, are also allowed to change. Moreover, shrinking the whole model space by model averaging or selection is a recommendable characteristic to solve the problem of over-fitting. Using alternative models that employ all the available indicators is not preferable due to the high dimensionality problem as well as the variation in predictors during boom and bust periods.

The empirical evidence that uses monthly returns on the S&P 500 price index is based on comparing the forecast performance of the DMA/DMS models to a number of alternative forecasting models which are classified as classical and Bayesian models, where Bayesian models include either time varying parameters, model uncertainty or both. Furthermore, DMA/DMS approaches provide evidence of which variables are more important in predicting stock market returns and give insight on the possibility of changing over time or not. The forecasted horizons are $h = 1$, $h = 3$, $h = 6$ and $h = 12$.

The key finding of this study is that applying the DMA/DMS approach leads to significant improvements in forecasting performance in comparison to other forecasting approaches. In particular, neither using only time varying parameter models nor relaxing parameter variation and changing models could improve the forecast performance. However, it is worth to note that there is model uncertainty rather than parameters instability in the case of predicting US stock market returns. Furthermore, the performance of the DMA/DMS models are better when including oil prices within predictors which denotes that oil prices are good predictors for stock market returns. The maximum number of predictors is changing over time and across horizons. Moreover, the posterior inclusion probabilities indicate that every predictor is important in the DMA forecasting at some times or some horizons but none is consistently the best.

Table 1: **Description Statistics: Monthly data**

Variable	Mean	Standard deviation
SR	0.517	3.522
IP	0.222	0.844
TSP	1.485	1.223
PI	0.307	0.348
M1	0.455	0.693
M2	0.552	0.355
FFR	-0.004	0.526
UE	-0.007	0.511
WTI	0.406	7.288
DY	-1.063	0.416
EP	-0.326	0.458
BM	0.515	0.254

Notes: SR = stock return, IP = industrial production growth, TS = term spread, PI = inflation rate, M1 = narrow money growth, M2 = broad money growth, FFR = interest rate, UE = change in unemployment rate, WTI = real oil price returns, DY = dividend yield, EP = earnings price ratio and BM = book to market ratio.

Table 2: **Unit root tests**

Variable	ADF	KPSS	PPerron
SR	-10.003*** (6)	0.139	-20.601*** (6)
IP	-9.685*** (3)	0.156	-7.995*** (6)
TSP	-4.425*** (4)	0.054	-4.529*** (6)
PI	-7.087*** (4)	0.646**	-13.159*** (6)
M1	-25.789*** (0)	0.301	-25.789*** (6)
M2	-26.067 (0)	0.107	-26.077*** (6)
FFR	-7.492*** (8)	0.089	-16.909*** (6)
UE	-7.433*** (5)	0.142	-21.821*** (6)
WTI	-11.492*** (5)	0.082	-20.552*** (6)
DY	-10.212*** (5)	0.111	20.244*** (6)
EP	-9.008*** (8)	0.036	-14.785*** (6)
BM	-7.404*** (8)	0.0921	-16.713*** (6)

Notes: SR = stock return, IP = industrial production growth, TS = term spread, PI = inflation rate, M1 = narrow money growth, M2 = broad money growth, IR = interest rate, UE = change in unemployment rate, WTI = oil price returns, DY = dividend yield, EP = earnings price ratio and BM = book to market ratio. ADF and PPerron are the Augmented Dickey Fuller and Phillips-Perron tests with the null hypotheses of unit root and KPSS test is based on a null hypothesis of stationary time series. *** and ** denote statistical significance at 1% and 5%, respectively.

Table 3: **MSFE for out-of-sample forecasts for Stock Returns**

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Clasic Models				
RW	3.870	3.799	3.999	4.023
AR	2.975	2.967	3.034	2.991
SR	3.545	3.647	3.499	3.801
Bayesian Models				
TVP-AR(1) at $\lambda = 0.99$	2.919	2.494	1.948	2.039
TVP-SR at $\lambda = 0.99$	2.036	2.111	2.074	2.019
$\lambda = \alpha$ **= 1.0**				
BMA	1.970	1.900	2.090	2.008
BMS	1.953	1.951	2.193	2.071
λ **=1.0** α **= 0.99** (No TVP)				
DMA	1.526	1.708	1.690	1.889
DMS	1.509	1.665	1.571	1.850
$\lambda = \alpha$ **= 0.99**				
DMA	1.614	1.718	1.695	1.992
DMS	1.561	1.685	1.658	1.912
$\lambda = \alpha$ **= 0.99** (Best Model-excluding OP)				
DMA	1.801	1.891	1.909	2.000
DMS	1.789	1.887	1.807	1.998

Note: Table entries are the results obtained from a forecasting exercise where the target variable is stock returns. The upper row shows the number of forecast horizon. The models in the first column are: classic models including naive random walk (RW) model, first order autoregressive (AR) model, and a simple regression (SR) model and Bayesian models comprise time varying parameters first order autoregressive model (TVP-AR), time varying parameters simple regression (TVP-SR) model, Bayesian model averaging (BMA), Bayesian model selection (BMS), Dynamic model averaging (DMA), and Dynamic model selection (DMS).

Figure 1: Plot of crude oil and stock market prices historical data

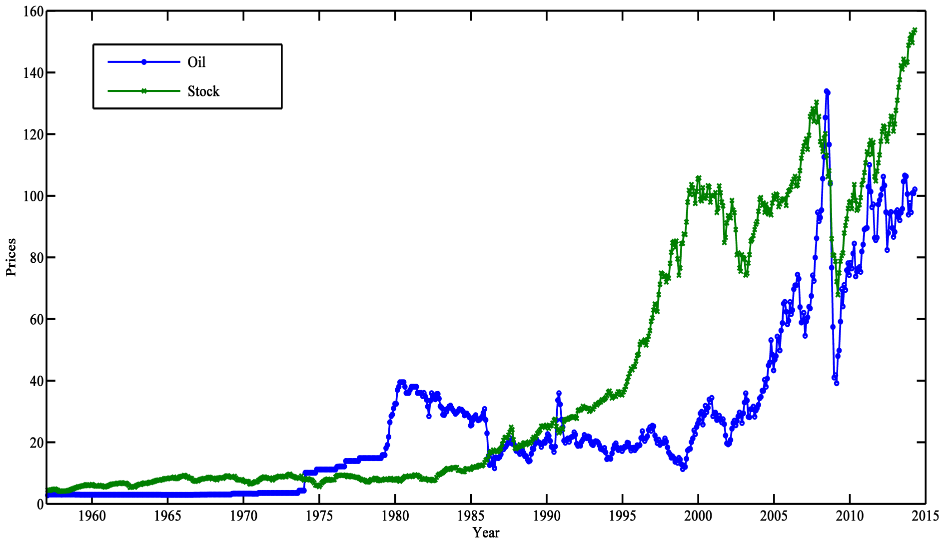
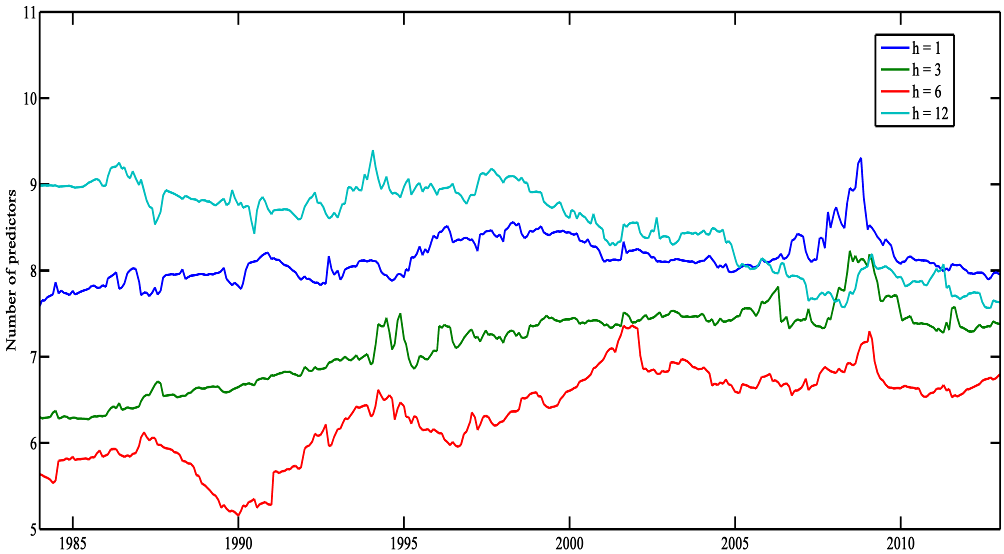
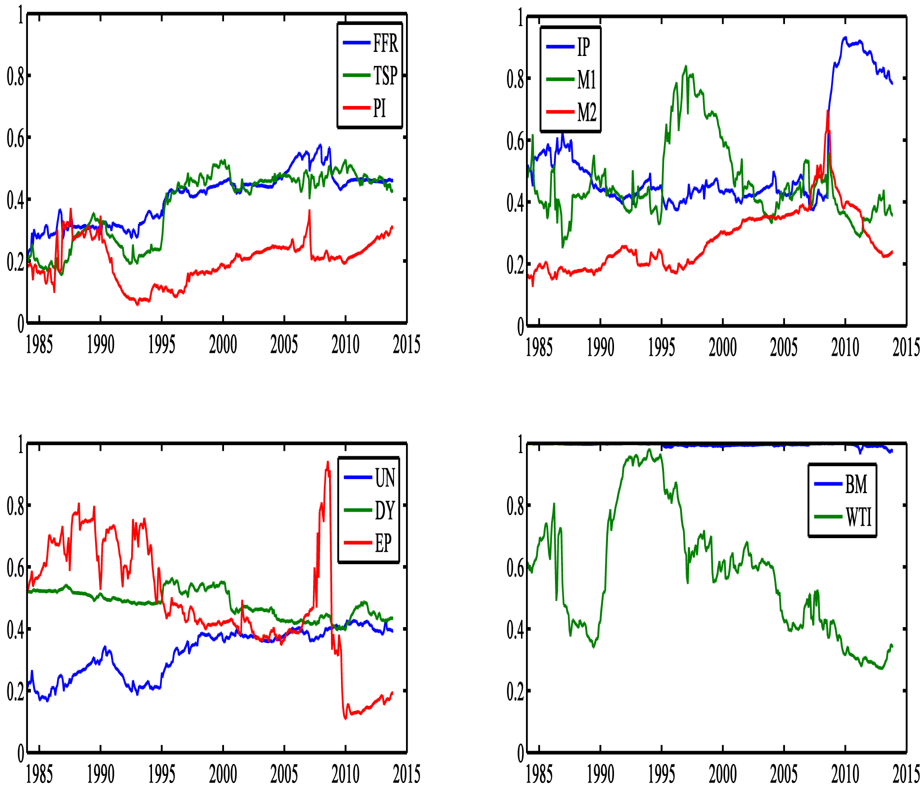


Figure 2: Expected number of predictors in each forecasting exercise



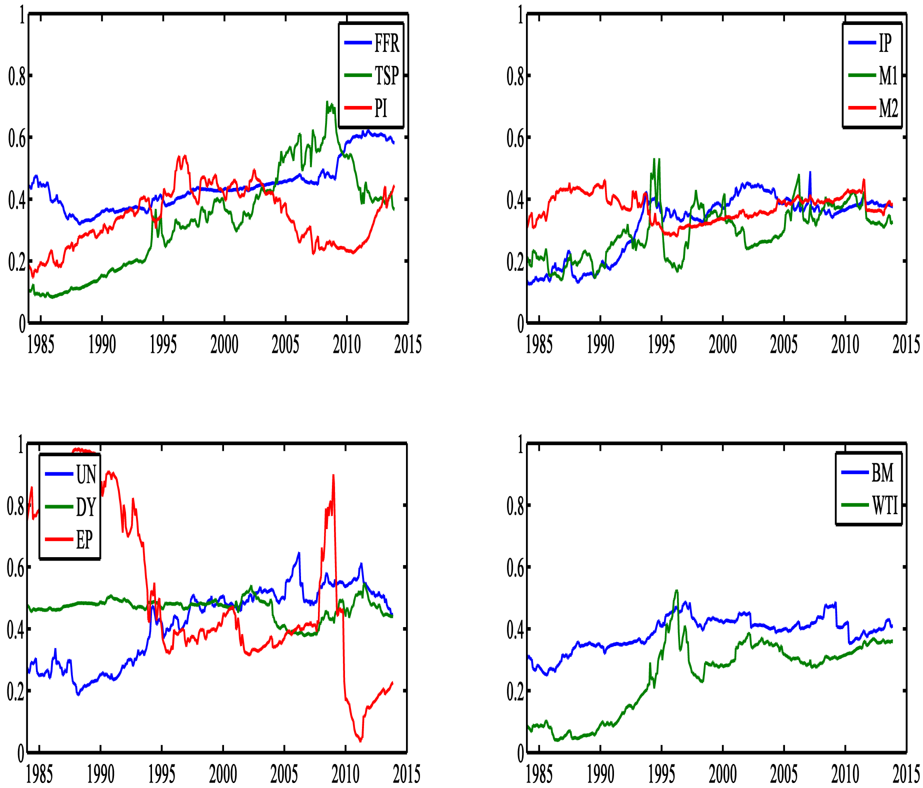
This figure provides information about the expected number of predictors used in each forecasting exercise, where h shows the number of forecasting steps. The considered forecast horizons are: one ($h = 1$), three ($h = 3$), six ($h = 6$), and twelve ($h = 12$).

Figure 3: Posterior inclusion probabilities of predictors (h = 1)



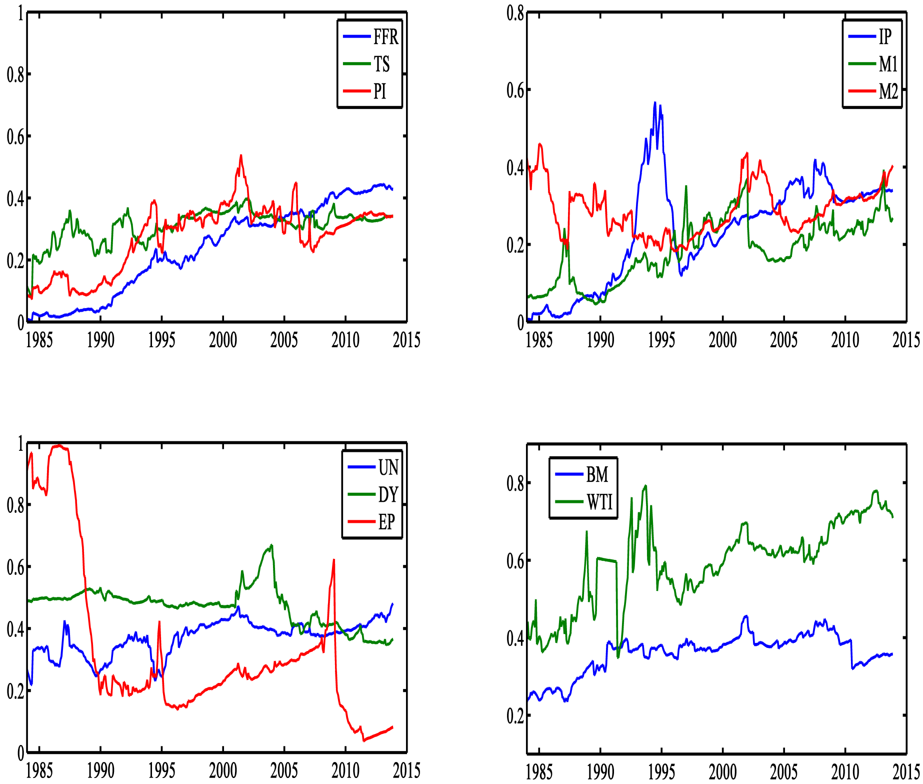
This figure shows the Posterior importance of each predictor at one ($h = 1$) step ahead forecasting exercise.

Figure 4: Posterior inclusion probabilities of predictors (h = 3)



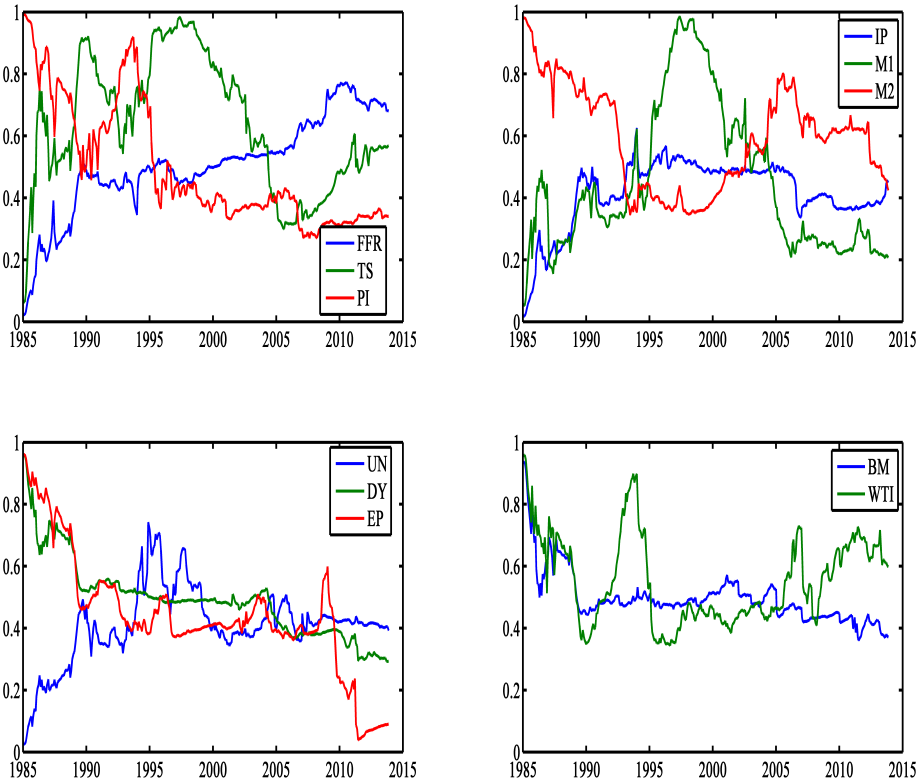
This figure shows the Posterior importance of each predictor at three ($h = 3$) steps ahead forecasting exercise.

Figure 5: Posterior inclusion probabilities of predictors (h = 6)



This figure shows the Posterior importance of each predictor at six ($h = 6$) steps ahead forecasting exercise.

Figure 6: Posterior inclusion probabilities of predictors (h = 12)



This figure shows the Posterior importance of each predictor at twelve ($h = 12$) steps ahead forecasting exercise.

References

Arouri, M. E. H., A. Lahiani, and D. K. Nguyen (2011). Return and volatility transmission between world oil prices and stock markets of the gcc countries. *Economic Modelling 28*(4), 1815–1825.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Basher, S. A., A. A. Haug, and P. Sadorsky (2012). Oil prices, exchange rates and emerging stock markets. *Energy Economics 34*(1), 227–240.

Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies 21*(4), 1509–1531.

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

Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance 33*(2), 211–223.

Chen, S.-S. (2010). Do higher oil prices push the stock market into bear territory? *Energy Economics 32*(2), 490–495.

Cogley, T. and T. Sargent (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics 8*(2), 262–302.

Cremers, K. M. (2002). Stock return predictability: A Bayesian model selection perspective. *Review of Financial Studies 15*(4), 1223–1249.

Cunado, J. and F. P. de Gracia (2014). Oil price shocks and stock market returns: Evidence for some European countries. *Energy Economics 42*, 365–377.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366a), 427–431.

Driesprong, G., B. Jacobsen, and B. Maat (2008). Striking oil: Another puzzle? *Journal of Financial Economics 89*(2), 307–327.

Grassi, S. and P. S. de Magistris (2015). It's all about volatility of volatility: Evidence from a two-factor stochastic volatility model. *Journal of Empirical Finance 30*, 62–78.

Groen, J. J., R. Paap, and F. Ravazzolo (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics 31*(1), 29–44.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hamilton, J. D. (2011). Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics 15*(S3), 364–378.

Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics 99*(3), 560–580.

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

IEA (2014). *Key world energy statistics.* International Energy Agency.

Jones, C. M. and G. Kaul (1996). Oil and the stock markets. *The Journal of Finance 51*(2), 463–491.

Justiniano, A. and G. E. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *The American Economic Review 98*(3), 604–641.

Kilian, L. and C. Park (2009). The impact of oil price shocks on the US stock market. *International Economic Review 50*(4), 1267–1287.

Kim, C.-J. and C. R. Nelson (1999). State-space models with regime switching: classical and Gibbs-sampling approaches with applications. *MIT Press Books 1.*

Koop, G. (2003). *Bayesian econometrics.* John Wiley and Sons.

Koop, G. and D. Korobilis (2011). UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling 28*(5), 2307–2318.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics 54*(1), 159–178.

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

Ludvigson, S. C. and S. Ng (2007). The empirical risk–return relation: a factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

Miller, J. I. and R. A. Ratti (2009). Crude oil and stock markets: Stability, instability, and bubbles. *Energy Economics 31*(4), 559–568.

Mollick, A. V. and T. A. Assefa (2013). US stock returns and oil prices: The tale from daily data and the 2008–2009 financial crisis. *Energy Economics 36*, 1–18.

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

Park, J. and R. A. Ratti (2008). Oil price shocks and stock markets in the US and 13 European countries. *Energy Economics 30*(5), 2587–2608.

Pesaran, M. H. and A. Timmermann (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance 50*(4), 1201–1228.

Pesaran, M. H. and A. Timmermann (2000). A recursive modelling approach to predicting UK stock returns. *The Economic Journal 110*(460), 159–191.

Pesaran, M. H. and A. Timmermann (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance 9*(5), 495–510.

Phillips, P. C. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*(2), 335–346.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics 52*(1), 52–66.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies 23*(2), 821–862.

Rapach, D. E., M. E. Wohar, and J. Rangvid (2005). Macro variables and international stock return predictability. *International Journal of Forecasting 21*(1), 137–166.

Rapach, D. E. and G. Zhou (2013). Forecasting stock returns. *Handbook of Economic Forecasting 2*(Part A), 328–383.

Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics 21*(5), 449–469.

Sarno, L. and G. Valente (2009). Exchange rates and fundamentals: Footloose or evolving relationship? *Journal of the European Economic Association 7*(4), 786–830.

Shiller, R. J. (2015). *Irrational exuberance.* Princeton University Press.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.