Predicting Daily Oil Prices: Linear and Non-Linear Models

Wassim Dbouk[1]
American University of Beirut

Ibrahim Jamali[2]
American University of Beirut

Abstract

In this paper, we assess the accuracy of linear and nonlinear models in predicting daily crude oil prices. Competing forecasts of crude oil prices are generated from parsimonious linear models which require no parameter estimation, as well as linear and nonlinear models. Two of the linear models that we employ exploit the informational content of oil demand and the increasing correlation between oil and equity prices and are novel to the literature. The nonlinear model that we consider is an artificial neural network. More specifically, we consider a bagged neural network, a neural network trained using the genetic algorithm as well as a neural network with fuzzy logic. We find that some of the linear models outperform the random walk in terms of out-of-sample statistical forecast accuracy. Our findings also suggest that while the buy-and-hold strategy dominates some of the models in terms of dollar payoffs and risk-adjusted returns under a *long-only* strategy, all the models that we consider generate higher dollar payoffs than the buy-and-hold strategy under the *short-only* strategy. An investor obtains the largest profits by trading based on the moving average convergence divergence which is a technical indicator.

Keywords: Forecasting, Crude Oil Market, Crude Oil Futures, Trading Strategy, Artificial Neural Network, Bootstrap Aggregation, Bagging, Genetic Algorithm, Fuzzy Logic, Error Correction Model, Transaction Costs, Autoregressive Distributed Lag, Financialization, Autoregressive Moving Average.

JEL Codes: G14, G17, Q41, Q47.

[1] Associate Professor., Department of Finance, Accounting and Managerial Economics, Olayan School of Business, American University of Beirut, Beirut 1107 2020, P.O. Box 11-0236, Riad El-Solh Street, Lebanon. Email: wd08@aub.edu.lb. Fax: +961-1-750 214. Tel: +961-1-340 460 (ext. 3762).

[2] Associate Professor., Department of Finance, Accounting and Managerial Economics, Olayan School of Business, American University of Beirut, Beirut 1107 2020, P.O. Box 11-0236, Riad El-Solh Street, Lebanon. Email: wd08@aub.edu.lb. Fax: +961-1-750 214. Tel: +961-1-340 460 (ext. 3770).

1. Introduction

Fluctuations in the nominal price of crude oil are known to affect the level of economic activity and consumer sentiment (Hamilton, 2009). The information contained in the term structure of crude oil futures markets has also been recently shown to predict inflation (Gospodinov, 2016) and to affect inflation expectations and breakeven inflation (Chen, 2009; Gospodinov and Wei, 2016). Research on the relation between fluctuations in the price of oil and economic activity resulted in a sizeable number of influential studies which document the adverse effects of oil price shocks on the level of economic activity (Hamilton, 1983, 1985, 2009, 2013; Hamilton and Herrera, 2004; Kilian, 2014). A parallel strand of the literature thoroughly investigates nonlinearities in the relationship between changes in crude oil prices and economic growth (Hamilton, 2003; 2011; Kilian and Vigfusson, 2011a, 2011b; 2013).[3]

Despite the paramount importance of the fluctuations in the price of oil on economic activity, predicting the price of oil continues to be a daunting task. This is unfortunate given that accurate forecasts of the price of oil are of central policy-making, trading and practical importance. From a policy-making perspective, accurate forecasts of the price of oil are helpful to central banks in designing policies that limit their impact on economic activity or mitigate their "pass-through" to inflation. In fact, under an inflation targeting monetary policy framework, central banks closely monitor the price of oil and seek to determine the effects of oil price fluctuations on their inflation target. The macroeconomic and policy-making importance of accurate oil price forecasts are clearly articulated in Alquist, Kilian and Vigfusson (2013): "central banks and private sector forecasters view the price of oil as one of the key variables in generating macroeconomic projections and in assessing macroeconomic risks".

From a trading and practical perspective, and given the existence of a highly liquid market for crude oil futures, accurate forecasts of the price of oil can be used by investors to implement trading strategies that potentially yield large (risk-adjusted) returns. The practical and policy-making importance of out-of-sample forecasts of oil prices has led to renewed interest in devising models which produce superior out-of-sample predictive accuracy.

While several models with good predictive power emerge from existing studies, this sizeable literature continues to highlight the inherent challenges in predicting the price of oil. The

[3] Yet another line of studies examines the relationship between the prices of shale oil or refined oil products and the price of crude oil (Kilian, 2010; 2016 among others).

resurgence of interest in predicting the price of oil was, at least in part, instigated by the influential study of Alquist, Kilian and Vigfusson (2013) which shows that the U.S. Energy Information Administration's forecasts of crude oil prices do not outperform a simple naïve forecast. However, subsequent work proposes several methods for generating forecasts which outperform the naïve (no change) forecast. Baumeister, Kilian and Zhou (2016) show that some refined oil product spreads are useful in predicting the price of oil at horizons up to two years. Baumeister and Kilian (2012) show that a recursive vector autoregressive models with global activity and oil market variables succeed in producing real-time forecasts of the real price of oil which outperform, in terms of directional accuracy, other time series forecasts as well as the no change forecast. In a similar vein, Baumeister and Kilian (2015) provide evidence that real-time forecasts of the price of oil which exploit information in the oil and refined oil markets or effectively combine various forecasts outperform the no-change forecast. Using mixed-data-frequency methods, Baumeister, Guérin and Kilian (2015) investigate the usefulness of incorporating high-frequency financial asset prices in predicting the price of oil. The authors find that their preferred mixed frequency model significantly improves the statistical accuracy of oil price forecasts vis-à-vis the no-change forecast.

In this paper, we assess the accuracy of linear and nonlinear models in predicting daily crude oil prices. Competing forecasts of crude oil prices are generated from parsimonious linear models that require no parameter estimation, an autoregressive moving average model, an autoregressive distributed lag model of crude oil prices and the Baltic dry Index, an error correction model of crude oil spot and futures prices, a demand model and artificial neural networks. While some of the latter models have already been used in the existing literature, we propose two models that are novel to the literature. The first is an adaptation of the demand model employed by academics and policy-makers (Hamilton, 2015; Bernanke, 2016) to understand the drivers of the recent decline in oil prices. The basic premise of the demand model is to employ the informational content of copper prices as a proxy of global economic activity (or oil demand). In view of the increasing correlation between oil and equity prices as well as the predictive power of exchange rates for commodity prices (Chen, Rogoff and Rossi, 2010), we also introduce a new model, referred to as the financialization model, which exploits the information content of equity prices and exchange rates. We adapt the autoregressive distributed lag model to predicting crude oil prices by incorporating changes in the Baltic dry index as an exogenous model. The latter model exploits the predictive content of the Baltic dry index which is widely viewed as a proxy of global economic activity. The nonlinear model that we consider is an artificial neural network. More specifically, we thoroughly examine the predictive accuracy of ANNs by generating forecasts from a two-layer feedforward network using three approaches. The first involves using bootstrap aggregation (or bagging) with the backpropagation algorithm. The second consists of training the network using the genetic algorithm while the third involves using fuzzy logic with bagging to generate forecasts.

As noted before, we predict daily crude oil prices. The use of daily data carries both advantages and disadvantages. On the one hand, using daily data ensures more accurate size and higher power for the statistical tests that we conduct (Alquist, Kilian, Vigfusson, 2013). On the other hand, daily prices (or price changes) are more volatile and are noisier than monthly prices. As a result, they are more difficult to predict. The trading importance of oil price forecasts motivates, to a large extent, our decision to employ daily data.

Our paper contributes to the literature on predicting the price of oil along several lines. To the best of our knowledge, this is the first paper to provide daily forecasts of the price of oil. In addition, we introduce and assess the predictive ability of two novel models: the demand and financialization models. In order to assess the economic significance of our results, we examine the (risk-adjusted) profitability of trading based on the directional forecast of the price of oil from the competing models. The profitability of trading based on the various model forecasts is compared to that of the Moving Average Convergence Divergence (MACD), a technical indicator, and the buy-and-hold strategies.

We find that some of the linear models outperform the random walk in terms of out-of-sample statistical forecast accuracy. Our findings also suggest that while the buy-and-hold strategy dominates some of the models in terms of dollar payoffs and risk-adjusted returns under a *long-only* strategy, all the models that we consider generate higher dollar payoffs than the buy-and-hold strategy under the *short-only* strategy. An investor obtains the largest profits by trading based on the MACD.

The outline of the paper is as follows. Section 2 begins with a discussion of the data and variables which we employ. Section 3 proceeds to introducing our forecasting models while Section 4 discusses the statistical forecast evaluation criteria we employ. In Section 5, we examine the economic significance of our forecasting results. Section 6 offers concluding remarks.

2. Data and Variables

2.1. Crude Oil Market and Futures

Crude oil is the world's most actively traded commodity. In fact, crude oil is an important input to production and several refined products, such as gasoline and diesel fuel can be obtained it. While derivative instruments on different types of oil are traded worldwide, crude oil futures, which trade on the New York Mercantile Exchange (NYMEX) of the Chicago Mercantile Exchange (CME), are the most widely used and liquid commodity futures contracts (Burghardt, 2008). Crude oil futures are settled based on the price of the West Texas Intermediate (WTI) oil delivered to Cushing, Oklahoma.

Crude oil futures are cash settled (i.e. marked to market daily).[4] The expiration cycle of the oil futures contract is monthly. That is, contracts expiring every month of the year, from January to December, are listed on the CME. Consecutive month contracts are listed for the current year as well as the next five years. The CME lists only the June and December contracts beyond the fifth year.[5] Crude oil futures trade virtually round the clock via the CME's electronic platform for six days of the week.[6] However, we employ in our empirical analysis data only from the regular trading hours. More specifically, we conduct the empirical analysis using the settlement price on the trading days.

In our empirical analysis, we forecast the changes in the price of the first (or nearest) futures contract. This choice is motivated by several considerations: First, the nearest futures contract is the most liquid one. Second, existing academic studies (Fama and French, 1987; 1988; Gospodinov and Ng, 2013) employ the price of the nearest futures contract in lieu of cash prices. In fact, Fama and French (1987, 1988) opt to rely on futures prices due to the lack of acurate spot price data. Third, Alquist, Kilian and Vigfusson (2013) note that central banks and international organizations routinely use the nearest oil futures price as a proxy for the spot price of oil.[7] Our use of futures contracts also stems from important trading considerations: Opening and maintaining a futures position involves lower transaction costs and is simpler to implement than transacting in the cash/spot market.

---

[4] The contract size is 1000 barrels and the minimum tick is $0.01.
[5] http://www.cmegroup.com/trading/energy/crude-oil/light-sweet-crude_contract_specifications.html?optid=425
[6] http://www.cmegroup.com/trading/why-futures/welcome-to-nymex-wti-light-sweet-crude-oil-futures.html
[7] Our forecasting results (available from the authors) are robust to using the cash price instead of the nearest futures price.

Crude oil futures price data are obtained from Datastream. Following the literature (Bessembinder, 1992; de Roon, Nijman and Veld, 2000; Gorton and Rouwenhorst, 2006), we employ a roll-over strategy to construct a continuous futures price series. The strategy consists of rolling over from the nearest to the next-to-nearest contract on the first day of the expiration month.[8]

2.2. Crude Oil Price Changes and Forecast Generation

In order to forecast from the linear and non-linear models introduced below, we divide our sample into an in-sample estimation (or training) and out-of-sample forecasting (or test) periods. We delineate our sample as follows: Our in-sample estimation (or training period when dealing with non-linear models) runs from January 2, 1990 to January 1, 2010 while our out-of-sample forecast evaluation (or test) sample spans the period January 4, 2010 to July 21, 2017. Note that our out-of-sample evaluation period is particularly challenging given that oil prices experienced booms and busts. For instance, a run-up in oil prices in 2007 was followed by a bust in 2008. Oil prices also exhibited another steep decline in 2014. We generate one-step-ahead forecasts from the competing models using a rolling (or sliding) window of 5217 trading days.

Let S_t denote the settlement price of the nearest oil futures contract. We test the null of a unit root in the crude oil futures prices, S_t, using an Augmented Dickey Fuller (ADF) test.[9] The ADF test results, provided in Panel A of Table 1, suggest that the null of a unit root cannot be rejected.[10] Let $s_t = \ln(S_t)$ denote the logarithm of settlement price of the nearest crude oil futures contract.[11] The unit root test results, reported in Panel B of Table 1, also show that the null of a unit root in the log oil price cannot be rejected. Therefore, when using linear (econometric) models to forecast

[8] That is, data from the expiration month are not used. This roll-over strategy allows us to avoid contract expiration effects (i.e. the high volatility near the contract expiration date as discussed in Bessembinder, Coughenour, Seguin and Monroe Smoller, 1996). Rolling over futures contracts also mimics traders' actual behavior. In fact, traders avoid taking delivery of the underlying commodity by closing out an open position prior to the contract's expiration (usually around the first notice day).

[9] We include an intercept and a trend in the ADF test equation. The number of lags is selected using the Bayesian Information Criterion (BIC).

[10] Given the lower power of ADF test against near-unit root alternatives, we also employ the the ADF with GLS detrending test of Elliott, Rothenberg and Stock (1996) which has excellent power properties as shown in the unit root literature. The results, reported in Panel A of Table 1, also suggest the presence of a unit root in crude oil prices.

[11] Again, we note that using cash (or spot) prices of crude oil does not materially affect our forecasting results. The results with the spot price of oil are available from the authors.

the price of oil, we avoid spurious results by predicting changes in the price of oil.[12] Changes in the price of oil are given by:

$$\Delta s_t = s_t - s_{t-1}. \tag{1}$$

Panel C of Table 1 provides the summary statistics of the crude oil price changes for the full sample as well as the in-sample and out-of-sample periods. Consistent with the stylized features of financial returns, the descriptive statistics show that the distribution of crude oil futures price changes is leptokurtic (i.e. exhibits excess kurtosis and fat tails) and exhibits some negative skewness. A Bera and Jarque (1981) test of normality (available from the authors) strongly rejects the null of normality in the crude oil returns. In contrast to the oil prices, the null of a unit root is strongly rejected for crude oil price changes.

[Insert Table 1 here]

While some of our parsimonious linear models straightforwardly generate one-step-ahead forecasts of S_t, which we denote as $\hat{S}_{t+1}$, we produce one-step-ahead forecasts of Δs_t from our econometric models so as to avoid spurious regression results. The forecasts of Δs_t, denoted as $\Delta \hat{s}_{t+1}$, are accumulated to obtain forecasts of the logarithm of the price of oil. Forecasts of the price of oil, $\hat{S}_{t+1}$, are, in turn, obtained using exponential values as $\hat{S}_{t+1} = \exp(\hat{s}_{t+1})$.

2.3. Other Predictors of Crude Oil Prices

The existing literature provides empirical evidence that several variables exhibit predictive power in forecasting crude oil prices or, more generally, commodity prices. The first predictor that we employ is the logarithmic change in the US dollar trade-weighted exchange rate. The use of this predictor is motivated by Chen, Rogoff and Rossi (2013)'s findings which suggest that exchange rates possess predictive power for commodity prices.

Our second predictor are the changes in a broad index of spot commodity prices. Alquist, Kilian and Vigfusson (2013) use this variable to predict crude oil prices. We employ the Commodity Research Bureau (CRB)'s spot commodity price index as our gauge of spot commodity prices. The other predictors that we employ are: Changes in the logarithm of the settlement price of the nearest copper futures contract, the change in the yield on the ten year U.S. Treasury note as well

[12] A forecast of the price of oil can be straightforwardly obtained by cumulating the forecasted changes in the price of oil.

as the Chicago Board Option Exchange (CBOE)'s option-implied volatility index, the VIX. Building on the observations of Hamilton (2015) and Bernanke (2016), we employ these latter variables as predictors in an oil demand model.

Our last predictor are the three month logarithmic changes in the Baltic Dry Index (BDI). Computed and disseminated by the Baltic Dry Exchange since May 1985, the BDI measures the cost of shipping major raw materials by sea (Schinas, Grau and Johns, 2014). Bakshi, Panayotov and Skoulakis (2011) provide empirical evidence that changes in the BDI are a useful predictor of commodity and equity returns as well as global economic activity.[13] Following Bakshi, Panayotov and Skoulakis (2011), we employ three-month logarithmic changes in the BDI as a predictor of crude oil price changes.[14]

Daily data on the U.S. dollar trade-weighted exchange rate the ten year U.S. Treasury note yield are obtained from the Federal Reserve Economic Data (FRED ©) database of the St. Louis Federal Reserve Bank.[15] Daily data on the CRB spot commodity price index, the VIX, the nearest copper futures settlement prices and the BDI are obtained from Datastream. Daily data on the VIX are available only starting January 2, 1990 thereby restricting the starting date of our sample.

Table 2 provides the summary statistics and cross-correlations of our predictors for the full sample.

[Insert Table 2 here]

Consistent with the observations of Bakshi, Panayotov and Skoulakis (2011), the descriptive statistics in Panel A of Table 2 show that changes in the BDI exhibit high volatility. While our predictors generally exhibit little first-order autocorrelation, the changes in the BDI are highly persistent. The cross-correlations, reported in Panel B of Table 2, indicate that pairwise correlations between our predictors are low.[16]

3. Linear and Nonlinear Forecasting Models

We discuss next the linear and nonlinear forecasting models that we employ to generate crude oil

[13] The BDI and Kilian's (2009) index of global real economic activity are constructed, as noted in Alquist, Kilian and Vigfusson (2013), using identical nominal data.
[14] Bakshi, Panayotov and Skoulakis (2011) note that three month changes in the BDI exhibit lower volatility than one period changes. Our own summary statistics corroborate this observation with daily data.
[15] The database can be found at: https://fred.stlouisfed.org/
[16] The highest correlation coefficient of 0.461 is between the changes in the CRB spot commodity prices and the changes in copper prices.

price forecasts. While some of these models have been previously employed in the oil forecasting literature, we introduce two novel models: The oil demand and financialization models. We also adapt the autoregressive distributed lag model to exploit the predictive content of the Baltic dry index.

3.1. The Random Walk Model

The first model we employ is a random walk (with no drift) model of crude oil prices. The random walk model is a widely used benchmark in the crude oil prediction literature (see, for example, Baumeister and Kilian, 2012, 2015; Alquist, Kilian and Vigfusson 2013). The model implies, in essence, a no-change forecast for the price of oil:

$$\hat{S}_{t+1} = S_t, \tag{2}$$

where $\hat{S}_{t+1}$ is the predicted oil futures price on day *t*+1. Consistent with the existing literature, we start by evaluating the predictive (and trading) performance of the random walk model. We henceforth refer to the random walk forecast as RANDOM WALK.

3.2. Parsimonious Linear Forecasting Models

Alquist, Kilian and Vigfusson (2013) propose several parsimonious models, which require no parameter estimation, for forecasting the price of oil. We next employ two of these models. The first model builds on Chen, Rogoff and Rossi (2010)'s empirical findings which suggest that exchange rates predict oil prices to propose the following parsimonious forecasting model:

$$\hat{S}_{t+1} = S_t(1+\Delta e_t), \tag{3}$$

where e_t is the logarithm of the trade-weighted U.S. dollar exchange rate. We assess the predictive accuracy of this simple parsimonious model. We refer to the forecasts from equation (3) as LINEAR 1.

The second parsimonious model we employ exploits the informational content of aggregate commodity prices in terms of out-of-sample forecasting. This model is given by:

$$\hat{S}_{t+1} = S_t(1+\Delta p_t^{com}), \tag{4}$$

where p_t^{com} is the natural logarithm of the level of the CRB commodity index.

The forecasts from equation (4) are referred to as LINEAR 2. Given that the models in equations (3) and (4) require no parameter estimation, they are not subject to estimation uncertainty.

3.3. The Oil Demand and Financialization Models

Several oil market observers and academics (Hamilton, 2015) have expressed the view that the recent decline in the price of oil is driven by weaker demand from emerging markets. The financial press, researchers, market observers and policy-makers[17] have also pointed to an increase in the correlation between equity index levels and crude oil prices and argued that the increased correlation is a manifestation of a common demand factor driving both equity and commodity prices.[18] In fact, the recent precipitous decline in oil prices has been attributed by the financial press to a weaker demand for oil by emerging markets.

These latter arguments are consistent with recent evidence suggesting that the increasing financialization of commodity markets results in increasing correlations between commodity and equity prices (Gorton and Rouwenhorst, 2006; Tang and Xiong, 2012; Cheng and Xiong, 2014).[19,20] In light of the heightened correlation between the oil and equity markets, researchers have recently examined whether the inclusion of financial variables in oil forecasting models delivers improvements in forecast accuracy. In specific, Baumeister, Guérin and Kilian (2015) use a mixed-data-frequency approach to explore the usefulness of high-frequency asset prices in predicting the price of oil. The authors' findings suggest that the predictive gains that can be achieved from using high-frequency financial data are modest.

The first parsimonious model attempts to exploit the increasing correlation between oil and equity prices (Tang and Xiong, 2012) as well as metals prices role as leading indicators of global

---

[17] Former Federal Reserve Board chairman Ben Bernanke examines the correlation between crude oil prices and S&P 500 levels (Bernanke, 2016).

[18] Evidence of a common factor in commodity and equity markets, however, continues to be elusive in the asset pricing literature. See, for example, Skiadopoulos (2013).

[19] The role of speculation in driving commodity prices is also a hotly debated topic in the literature. While some researchers argue that the participation of commodity trading advisors and hedge funds in commodity markets might result in an increase in the correlation between commodity and equity prices (Buyuksahin and Robe, 2011; Singleton, 2013), the predominant view appears to be that fluctuations in commodity prices cannot be straightforwardly attributed to speculation or hedge fund participation, See, for example, Stoll and Whaley (2010), Fattouh, Kilian and Mahadeva (2013) and Irwin and Sanders (2012).

[20] Several articles in the financial press refer to the increased correlation between oil and equity prices and tie the heightened correlation to weaker demand for oil from emerging economies. See, for example, http://www.wsj.com/articles/oil-stocks-dance-the-bear-market-tango-1453722783

economic activity (Caldara, Cavallo and Iacovello, 2016; Pindyck and Rotemberg, 1990; Labys, Achouch and Terraza, 1999):

$$\Delta\hat{s}_{t+1} = a + b_1\Delta s_t^c + b_2 r_t^{MSCI} + e_{t+1}, \tag{5}$$

where r_t^{MSCI} are the continuously compounded returns on the emerging markets MSCI index and s_t^c is the logarithm of the settlement price of the nearest copper futures contract. As stated earlier, the price of copper serves as a proxy for global economic health. The forecasts from equation (5) are referred to as FINANCIALIZATION.

We also adapt the oil demand model of Hamilton (2015) and Bernanke (2016) to forecasting the price of oil. The model is given by:

$$\Delta\hat{s}_{t+1} = \alpha + \beta_1\Delta s_t^c + \beta_2\Delta e_t + \beta_3\Delta r_t^{10} + \beta_4 VIX_t + \varepsilon_{t+1}, \tag{6}$$

where s_t^c is the logarithm of the settlement price of the nearest copper futures contract, e_t is the logarithm of the trade-weighted US dollar exchange rate, r_t^{10} is the yield on the ten year U.S. Treasury note and VIX_t is the CBOE option-implied volatility index for the S&P 500. In equation (6), the price of copper again serves as a proxy for demand conditions in the global economy. Changes in the exchange rate are included in view of evidence suggesting their usefulness in predicting commodity prices (Chen, Rogoff and Rossi, 2010). The VIX index is widely viewed as an investor fear gauge in equity markets (Whaley, 2000) and can be employed to extract the variance risk premium (Bekaert and Hoerova, 2014, Bollerslev, Tauchen and Zhou, 2009). Given the VIX's importance as a measure of uncertainty and following Bernanke (2016), we include the changes in the VIX index as a predictor in the proposed demand model. The forecasts from the demand model in equation (5) are referred to hereafter as DEMAND.

3.4. Autoregressive Moving Average (ARMA) Model

Autoregressive moving average models (ARMA) are simple linear time series models that have been widely employed to predict crude oil price changes (Baumeister and Kilian, 2012), interest rates spreads/changes (Gospodinov and Jamali, 2011; Dbouk, Jamali and Kryzanowski, 2016) and other financial asset returns.

An ARMA (p,q) relates the change in the price of crude oil at time t+1 to p of its own lags as well as to q lags of a white noise error term:

$$\Delta\hat{s}_{t+1} = \phi_0 + \phi_1 \Delta s_t + \phi_2 \Delta s_{t-1} + ... + \phi_p \Delta s_{t-p+1} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-1} - ... - \theta_q u_{t-q+1}, \qquad (7)$$

where u_t is a white noise error term. The optimal autoregressive (AR) and moving average (MA) lag lengths, $p*$ and $q*$, are selected by searching over several possible combinations of p and q and selecting the model which minimizes the Akaike Information Criterion (AIC). The selected model is an ARMA (4,2) and we verify that the model's residuals do not exhibit any remaining serial correlation. The forecasts generated from the ARMA model in equation (7) are henceforth referred to as ARMA.

3.5. Error Correction Model (ECM)

Spot and futures prices should exhibit, according to economic theory, a long-run cointegrating relationship. Existing research (Gospodinov and Jamali, 2011; Brooks, Rew and Ritson, 2001) exploits the existence of such a cointegrating relationship for predictive purposes. For instance, Coppola (2008) provides empirical evidence that the out-of-sample predictive accuracy of a vector error correction model of crude oil spot and futures prices is superior to that of a random walk.

Based on these observations, we employ an ECM model of crude oil spot and futures prices to predict crude oil price changes. One advantage of the ECM is the model's ability to exploit the informational content of futures prices in predicting the price of oil. A second advantage of the ECM model is its ability to capture and exploit, in terms of out-of-sample forecasting, the long-run (or equilibrium) relationship between the spot and futures prices of crude oil. Again, we note that we proxy the spot price of oil using the nearest futures contract while we consider the futures price to be the next-to-nearest (or second) futures contract. Let $f_t^{(1)}$ denote the logarithm of the next-to-nearest crude oil futures prices.

We employ the Engle and Granger (1987) two-step approach to test for cointegration between s_t and $f_t^{(1)}$. In the first step, we estimate the cointegrating regression (i.e. cointegrating vector) using our in-sample observations:

$$s_t = \gamma_0 + \gamma_1 f_t^{(1)} + \zeta_t, \qquad (8)$$

and test the residuals of the above regression, $\hat{\zeta}_t$, for a unit root. Rejecting the null of a unit root in the residuals of the cointegrating regression in equation (8) implies that s_t and $f_t^{(1)}$ are cointegrated.[21]

Our in-sample estimation results for equation (8) yield an estimate of the intercept, $\hat{\gamma}_0$, of 0.01 and an estimate of the slope coefficient, $\hat{\gamma}_1$, of 0.99. Both coefficients are significant at the 1% level. The null of a unit root in the residuals, $\hat{\zeta}_t$, is rejected (using the ADF and the ADF-GLS tests) at the 1% and 5% levels, respectively. We view these results as suggestive of the presence of a cointegrating relationship between s_t and $f_t^{(1)}$ and, accordingly, proceed to estimate an ECM. As noted by Brooks, Rew and Ritson (2001), the Granger representation theorem implies that the variables can be modelled using an ECM. The ECM model is given by:

$$\Delta\hat{s}_{t+1} = \beta_0 + \delta\hat{\zeta}_t + \beta_1\Delta s_t + ... + \beta_p\Delta s_{t-p+1} + \alpha_1\Delta f_t^{(1)} + ... + \alpha_q\Delta f_{t-q+1}^{(1)} + v_{t+1}. \tag{9}$$

In equation (9), the error correction term is $\hat{\zeta}_t = s_t - \gamma_0 - \hat{\gamma}_1 f_t^{(1)}$ while the speed of adjustment parameter (to long-run equilibrium) is $\hat{\delta}$. Optimal lag lengths, *p** and *q**, of eight are selected by ensuring that the ECM's residuals do not exhibit any residual autocorrelation. The forecasts from the ECM model are henceforth referred to as ECM.

3.6. Autoregressive Distributed Lag (ARDL) Model

As discussed earlier, existing research provides empirical evidence that the changes in the BDI index are a useful predictor of crude oil prices. In fact, the BDI is closely scrutinized by the financial press and practitioners and the common view echoed in the financial press is that the index serves as a leading indicator of global economic and trade activity.[22]

---

[21] In order to have a cointegrating relationship between s_t and $f_t^{(1)}$, both variables should be integrated of the same order (i.e. should contain a unit root or are integrated of order one). We test for a unit root in the logarithm of the settlement price of the second oil futures contract and our conclusions indicate that the null of a unit root cannot be rejected.

[22] See, for example, www.economist.com/blogs/economist-explains/2015/03/economist-explains-7, hwww.newyorker.com/business/currency/the-surprising-relevance-of-the-baltic-dry-index and www.bloomberg.com/news/articles/2016-01-05/baltic-dry-ship-index-tumbles-to-fresh-record-amid-china-turmoil.

We assess the predictive power of the BDI index in forecasting crude oil prices using an Autoregressive Distributed Lag (ARDL) model. Let g_t denote the three month (logarithmic) change in the BDI. The ARDL (p,q) model is given by:

$$\Delta\hat{s}_{t+1} = \phi_0 + \phi_1\Delta s_t + ... + \phi_p\Delta s_{t-p+1} + \lambda_1 g_{t+1} + ... + \lambda_q g_{t-q} + \eta_{t+1}, \tag{10}$$

where the autoregressive lag order is p and the distributed lag order is q. The optimal lag lengths, p^* and q^*, are selected by searching over all possible combinations of lag orders up to twelve (i.e. up to $p = 12$ and $q = 12$) and selecting the model which minimizes the AIC.

Our selection procedure yields an ARDL (8,1) and the selected model is employed for out-of-sample forecasting. Forecasts from the ARDL model are referred to, henceforth, by ARDL.

3.7. Artificial Neural Networks: Bagging, Genetic Algorithm and Fuzzy Logic

The linear forecasting models introduced above do not allow for nonlinearities in oil price dynamics. Accounting for nonlinearities can possibly yield forecast improvements. In this section, Artificial Neural Networks (ANNs) are employed to capture nonlinearities in the price of crude oil and produce out-of-sample forecasts.[23] Neural networks are popular for modeling and forecasting financial and macroeconomic times series because of their ability to approximate nonlinear dynamics of unknown form arbitrarily closely (Franses and van Dijk, 2000).[24] More specifically, researchers have examined ANNs' predictive ability in the foreign exchange (Yu, Wang and Lai, 2010; Franses and van Griensven, 1998; Franses and van Homelen, 1998) equity (Refenes, Zaparanis and Francis, 1994; White, 1988), fixed income (Swanson and White, 1995) and commodity (Kohzadi *et al.*, 1996) markets.[25] Studies which explore the predictive ability of ANNs for macroeconomic variables include Maasoumi, Khotanzad and Abaye (1994), Swanson and White (1997) and Teräsvirta, van Dijk and Medeiros (2005).

Following Zhang, Putawo and Hu (1998), an ANN can be written as:

---

[23] The manuscript in general and this section, in particular, has benefited tremendously from the comments of anonymous reviewers.

[24] An introduction to neural networks from the perspective of econometricians is given in Kuan and White (1994).

[25] An exhaustive list of studies which employ ANN in financial forecasting is provided in Zhang, Patuwo and Hu (1998)'s authoritative review. Based on the vast number of studies reviewed in their survey, Zhang, Patuwo and Hu (1998) provide a comparative assessment of the predictive performance of ANNs relative to linear models. In addition to forecasting, ANNs have also been used in option pricing (Gradojevic, Kukolj and Gencay, 2009) and are popular in the bankruptcy prediction literature (see Zhang, Patuwo and Hu (1998) and the references therein). Because of their pattern recognition abilities, some researchers (see, for example, Kryznanowski, Galler and Wright, 1993) also explore ANNs' usefulness in stock selection.

$$\hat{S}_{t+1} = f(x_1, x_2, \ldots, x_p), \tag{11}$$

where $f(.)$ denotes the (nonlinear) functional relation between p independent variables, known as inputs and denoted by $x_1, \ldots, x_p$, and the price of oil, referred to as the output. In a time series forecasting setup, the inputs are typically lagged values of the series itself. In this paper, the vector of inputs includes only lagged values of S_t so that equation (11) specializes to $\hat{S}_{t+1} = f(S_t, S_{t-1}, \ldots, S_{t-p})$.

In their authoritative survey on forecasting with ANNs, Zhang, Putawo and Hu (1998) note that selecting the network's architecture is one of the important choices to be made by a researcher. Selecting the ANN's architecture involves choices pertaining to (i) the number of input nodes, (ii) the number of hidden layers and (iii) the number of output nodes.[26] The choice of hidden nodes is a particularly important, albeit challenging, decision that the forecaster should make. By increasing the number of nodes, a forecaster can approximate nonlinear dynamics of unknown form arbitrarily closely and can train the ANN to any desired level of in-sample fit. This, in turn, might lead to overfitting and does not necessarily yield improved out-of-sample performance.

We employ a two-layer feedforward network with three inputs (i.e. three lags of S_t). While we choose the number of inputs by trial-and-error,[27] the choice of two layers is motivated by Zhang (1994)'s findings which demonstrate that forecasts of a particle time series from the Santa Fe forecasting competition that are generated from a network with two hidden layers are superior to those from a network with a single layer.[28] We use the hyperbolic tangent (*tanh*) function as an activation function. Our starting point is to use the backpropagation algorithm, which is a gradient steepest descent optimization algorithm (Zhang, Putawo and Hu, 1998), to train the ANN. However, when the ANN is trained using the backpropagation mechanism, the forecasting results

[26] In our case, we only have single output, which is the price of oil. Other important choices to be made include selecting the activation function, the training algorithm and the training and test samples (Zhang, Putawo and Hu, 1998).

[27] The trial-and-error approach is referred to by Zhang, Putawo and Hu (1998) as the shotgun approach. The authors also discuss heuristics (or rules of thumb) relating to the choice of number of nodes. As discussed next, we follow Szafranek (2017) by experimenting with (i.e., randomizing) the number of nodes when we implement bootstrap aggregation. Out results suggest that the ANN with two hidden nodes and three inputs generates superior forecasts than when we randomize the number of inputs and layers under bootstrap aggregation.

[28] In addition, the choice of two hidden layers is consistent with Zhang, Patuwo and Hu (1998) who note that "In our view, one hidden layer may be enough for most forecasting problems. However, two hidden layers may give better results for some specific problems, especially when one hidden layer network is overladen with too many hidden nodes to give satisfactory results."

hinge upon the starting (guess) values and the algorithm might suffer from local optima (Zhang, Putawo and Hu, 1998). The high variance and instability of the forecasts from an ANN are a result of the training algorithm being stuck in such local optima. Therefore, relying on a single ANN will lead to inferior forecasts given that ANNs are unstable predictors and, as articulately discussed by Zhang, Putawo and Hu (1998), are characterized by high variance. We address the high variance of ANN forecasts using a three-pronged approach.

First, recent contributions to the literature resort to, in the parlance of artificial intelligence and machine learning researchers, ensembling neural networks (Zhou, Wu and Tang, 2002) in order to obtain more robust forecasts. That is, researchers build an ensemble of neural networks and aggregate their output to obtain a single prediction. The aggregate output typically exhibits higher predictive accuracy and lower variance than the forecast from a single ANN.

A popular approach to aggregating neural networks is bootstrap aggregation (also known as bagging).[29] Bagging consists of estimating ANNs using several bootstrap (training) samples and generating forecasts from each sample (Szafranek, 2017). The forecasts from bootstrap samples are then averaged to produce a single forecast which typically exhibits lower variance and higher prediction accuracy than a single ANN's forecast (Khwaja *et al*., 2015). When implementing bagging, we employ the Efron bootstrap and follow the literature in econometrics (Szafranek, 2017) by using one hundred replications.[30] The forecasts from the ANN with bagging are, henceforth, referred to as BANN.

Second, we train the neural network using the genetic algorithm. The genetic algorithm is an evolutionary search process which overcomes the problem of local minima by assisting the researcher in finding better guess (starting) values for the search algorithm (McNelis, 2005).[31] The forecasts that we generate from the neural network with the genetic algorithm as a training algorithm are referred to as GAANN.

Third, we combine fuzzy logic with neural networks to enhance the forecasting performance of the ANN. Fuzzy logic is a computational approach which permits partial truths instead of binary

[29] Bagging has also been shown to reduce the forecast errors of the models when applied to linear forecasting models by econometricians. See, for example, Inoue and Kilian (2008) and Rapach and Strauss (2010).
[30] Existing research (Breiman, 1996) suggests that the gain from bagging stabilizes after twenty-five replications. Nonetheless, we follow the econometrics literature (Inoue and Kilian, 2008; Rapach and Strauss 2010; Szafranek, 2017) by using one hundred bootstrap replications.
[31] McNelis (2005) identifies the following steps for the genetic algorithm: (i) population creation, (ii) selection, (iii) crossover, (iv) mutation, (v) election tournament, (vi) elitism and (vi) convergence.

signals (Novák, Perfilieva and Močkoř, 1999). As noted in Buckley and Hayashi (1994), a fuzzy neural network is one with fuzzy signals and/or fuzzy weights. We also apply bagging to the fuzzy neural network and the forecasts are referred to as FANN.

We guard against overfitting by (i) using a small number of nodes, (ii) implementing an early stopping mechanism and (iii) using a validation sample. In fact, Nakamura (2005) highlights the importance of early stopping in improving the predictive performance of a neural network. Our training sample extends from January 2, 1990 to January 1, 2008. The validation sample spans the period January 2, 2008 to January 1, 2010 while the test period is, consistent with the out-of-sample forecasting period for the linear models, January 4, 2010 to July 21, 2017.

4. Statistical Forecast Accuracy

We start by assessing the statistical forecast accuracy of our competing models. To do so, we employ a number of commonly used forecast accuracy measures. These criterions are: the Mean Absolute Error (*MAE*), the Root Mean Squared Error (*RMSE*) and the Mean Absolute Percentage Error (*MAPE*). Each of the latter three criterions uses a different loss function. That is, each of these criterions penalizes large forecast errors differently. In addition to the prior criteria, we employ Theil's *U* statistic and the out-of-sample R^2. We discuss each of these statistical forecast accuracy measures next.

4.1. Mean Absolute Error (MAE)

The Mean Absolute Error (*MAE*) uses the absolute value of the forecast errors as a loss function. Let T_f denote the size of out-of-sample forecast evaluation sample. The *MAE* is given by:

$$MAE = \frac{1}{T_f}\sum_{t=1}^{T_f}\left|S_t - \hat{S}_i^t\right|, \tag{12}$$

for $i = 1, \ldots, 11$ and where 1,..., T_f is the out-of-sample forecast sample and $\hat{S}_t$ is a forecast of the price of oil from one of the competing models. The model with the smallest *MAE* exhibits the highest predictive accuracy according to this criterion.

4.2. Root Mean Square Error (RMSE)

Unlike the *MAE*, the *RMSE* uses a quadratic loss function given by:

$$RMSE = \frac{1}{T_f}\sum_{t=1}^{T_f}\left(S_t - \hat{S}_t^{\,i}\right)^2 , \tag{13}$$

for $i = 1, \ldots, 11$. Again, the model with the lowest *RMSE* would outperform competing models in terms of out-of-sample forecasting. We note, however, that large forecast errors are penalized more severely under the *RMSE* than the *MAE*.

4.3. *Mean Absolute Percentage Error (MAPE)*

Another criterion that we employ to assess statistical forecast accuracy is the Mean Absolute Percentage Error (MAPE). *MAPE* is given by:

$$MAPE = \frac{100}{T_f}\sum_{t=1}^{T_f}\left|\frac{S_t - \hat{S}_t^{\,i}}{S_t}\right| , \tag{14}$$

for $i = 1, \ldots, 11$ There are two advantages to using *MAPE*: First, this criterion is bounded by zero from below. Second, *MAPE* can be interpreted as a percentage error.

4.4. *Theil's U Statistic*

Theil's *U* statistic is another widely used statistical forecast accuracy criterion. First proposed by Theil (1966), this statistic is given by:

$$U = \frac{\sqrt{\sum_{t=1}^{T_f}\left(\frac{S_t - \hat{S}_t^{\,i}}{S_t}\right)^2}}{\sqrt{\sum_{t=1}^{T_f}\left(\frac{S_t - \hat{S}_t^{RW}}{S_t}\right)^2}} , \tag{15}$$

for $i = 1, \ldots, 10$ and where $\hat{S}_t^{RW}$ is a forecast of the price of oil from the random walk benchmark.

A Theil's *U* statistic lower than one is indicative that a model's forecast is more accurate than the random walk while a value of the *U* statistic greater than one implies that the model under consideration is less accurate than the random walk. When Theil's *U* statistic is equal to one, the random walk and the model under consideration are equally accurate.

4.5. Out-of-Sample R^2 (OoS)

The Out-of-Sample R^2 (OoS) is an increasingly more popular method for assessing forecast accuracy in the financial economics and forecasting literatures.[32] The *OoS* R^2 is given by:

$$OoS = 1 - \frac{\sum_{t=1}^{T_f}\left(S_t - \hat{S}_t^i\right)^2}{\sum_{t=1}^{T_f}\left(S_t - \hat{S}_t^{RW}\right)^2} = 1 - \frac{MSE^i}{MSE^{RW}}, \tag{16}$$

for $i = 1, \ldots, 10$ and where $\hat{S}_t^{RW}$ is the forecast of crude oil prices from the random walk (benchmark) model. Note that *OoS* can be equivalently written in terms of the MSE of the model under consideration, *MSE*, and that of the random walk benchmark, MSE^{RW}.

A positive (negative) value of the *OoS* indicates that the model under consideration outperforms (underperforms) the random walk in terms of out-of-sample forecast accuracy.

4.6. The Diebold and Mariano (1995) and Clark and West (2007) Tests

While ranking the models based on the various loss function is a useful starting point, we also assess the statistical significance of differences in the loss functions. First, we employ the modified Diebold and Mariano (1995), henceforth DM, test. Denote the difference in loss functions by $d_{t+1} \equiv \Delta L_{t+1} = L(S_{t+1} - \hat{S}_{t+1}) - L(S_{t+1} - \hat{S}_{t+1}^R)$ where $\hat{S}_{t+1}$ is the forecast from one the competing and $\hat{S}_{t+1}^{RW}$ is the forecast from the random walk model.

The DM test for the null of equal predictive accuracy $H_0 : E[d_{t+1}] = 0$ is given by the test statistic $\bar{d}/\hat{\sigma}_{\bar{d}}$ where $\bar{d}$ is the mean sample differential and $\hat{\sigma}_{\bar{d}}$ is a consistent estimate of the standard deviation of $\bar{d}$. Under certain regularity conditions, the latter test statistic follows the standard normal distribution. Harvey, Leybourne and Newbold (1997) introduce a degrees of freedom adjustment for the variance of the DM statistic and employ the critical values from the *t* distribution. We employ the Harvey, Leybourne and Newbold (1997) modified version of the DM test and refer to it as MDM.

When comparing nested models, the DM statistic follows a nonstandard distribution. Therefore, existing studies (Rapach, Strauss and Zhou, 2010 among others) employ the Clark and West (2007) modified version of the DM statistic. The Clark and West (2007) *MSPE-adjusted* statistic yields

[32] The *OoS* is used, among others, by Welch and Goyal (2008) and Campbell and Thompson (2008).

asymptotically valid comparisons (with the standard normal critical values) for nested linear models. The *MSPE-adjusted* statistic can be computed by first defining:

$$f_{t+1} = (\Delta S_{t+1} - \Delta \hat{S}_{t+1}^{RW})^2 - [(\Delta S_{t+1} - \Delta \hat{S}_{t+1})^2 - (\Delta \hat{S}_{t+1}^{RW} - \Delta \hat{S}_{t+1})^2]. \tag{17}$$

The statistic can then be obtained by regressing f_{t+1} on a constant and computing the t-statistic and the p-value for a one-sided upper tail test (Rapach, Strauss and Zhou, 2010).

4.7. Conditional Test of Forecasting Performance

We provide a more detailed assessment of the predictive performance of the different models by also employing the conditional test of predictive accuracy of Giacomini and White (2006). The Giacomini and White (2006) test is given by the Generalized Method of Moments (GMM) quadratic form:

$$GW = T_f \left(T_f^{-1} \sum_{t=0}^{T_f} v_{t+1} \Delta L_{t+1} \right)' W^{-1} \left(T_f^{-1} \sum_{t=0}^{T_f} v_{t+1} \Delta L_{t+1} \right), \tag{18}$$

where W is the optimal weighting matrix in the GMM problem. The GW statistic follows a χ^2 distribution. In our forecasting exercise, the GW statistic follows a χ^2 distribution with two degrees of freedom. A (negative) positive value of the statistic is indicative that the competing model (under) outperforms the random walk.

The null hypothesis of the GW test is one of equal predictive accuracy between the random walk benchmark and the competing model's forecast. The GW test has several advantages (Elliott and Timmermann, 2016). First, it accounts for the effects of estimation error on the forecasts. Second, it is a finite-sample test, as opposed to an asymptotic one which compares the predictive performance of two models in the population, so that the estimation error does not vanish (Elliott and Timmermann, 2016).[33]

4.8. Statistical Forecast Accuracy Results

Panel A of Table 3 provides the *MAE*, *RMSE*, *MAPE*, Theil *U*, *OoS* statistics for our competing models.

[Insert Table 3 here]

---

[33] The Diebold and Mariano (1995) test, which can also be used to assess the significance of the differences in the loss functions among the competing forecasts cannot be used for comparing nested models.

The results indicate that the financialization and demand models outperform all the competing models in terms of MAE, RMSE and MAPE. The third best performing model according to the MAE is the random walk while two models (the ECM and LINEAR 1 models) outperform the random walk according to the RMSE. While the FANN's predictive accuracy is broadly in line with that of the linear models, the BANN performs significantly worse than the linear models. This is likely to be a result of the backpropagation algorithm being stuck in local minima. The GAANN's statistical accuracy, while worse than that of the linear models, is closer to that of the FANN than that of the BANN. Similar conclusions regarding the superior out-of-sample forecast accuracy of the financialization and demand models can be gleaned from Theil *U*'s statistic. Namely, the Theil *U* statistic is less than one, albeit only marginally, for the demand, financialization, LINEAR 1 and ECM models again suggesting that these models outperform the random walk.

Panel B of Table 3 reports the OoS R^2 statistic, the Giacomini and White (2006) test, the modified Diebold and Mariano (1995) test as well as the *MSPE-adjusted* statistic of Clark and West (2007). The OoS R^2 is positive for the demand, financialization, LINEAR 1 and ECM models while it is negative for all the remaining models. When the GW test is used to assess predictive accuracy, our results indicate that the random walk benchmark statistically significantly (at the 5% level) outperforms all the models except the demand, ARMA, ECM, financialization and FANN models. For the latter five models, the null of equal predictive accuracy between the model and the random walk cannot be rejected.[34] According to the MDM statistic, only the ECM model performs better than the random walk albeit the over performance is not statistically significant. Interestingly, the results of the Clark and West (2007) *MSPE-adjusted* statistic, which is better suited for comparing nested models, suggest that the ECM and FANN statistically significantly outperform the random walk.

Overall, two main findings emerge from our results. First, outperforming the random walk benchmark in terms of forecasting daily crude oil prices is a very challenging task. Second, there is some evidence that the demand, financialization, ECM and FANN models appear to provide, out-of-sample predictive gains vis-à-vis the random walk. We turn next to assessing the economic significance of the uncovered predictability.

[34] Note that the 5% level of significance is employed in light of the large out-of-sample forecasting sample. The GW test results also show that the LINEAR 1 model is outperformed by the random walk.

4.9. Forecast Optimality, Forecast Encompassing and Forecast Combination

A desirable property of forecasts is that of optimality with respect to an information set (Diebold, 2014). Optimality necessitates that forecast errors, $S_t - \hat{S}_t$, be unpredictable given the information at time *t*. Instead of testing predictability in the forecast errors directly, researchers commonly employ the Mincer and Zarnowitz (1969) regression to test for forecast optimality. The Mincer-Zarnowitz regression is given by:

$$S_t = \beta_0 + \beta_1 . \hat{S}_t + e_t, \qquad (19)$$

for $t = 1, \ldots, T_f$. Forecast optimality entails that the null hypothesis $H_0 : \beta_0 = 0, \beta_1 = 1$ is not rejected. We next assess the optimality of the forecasts from each of our competing models by estimating the Mincer-Zarnowitz regression in equation (19) for each of our forecasts.

The results from the forecast optimality tests in equation (19) are provided in Panel A of Table 4.

[Insert Table 4 here]

Our results indicate that the null of forecast optimality cannot be rejected, at the 5% level, for any of forecasts except for the BANN and FANN.[35]

The existing literature emphasizes the predictive gains that can be achieved by combining forecasts (Timmermann, 2006). These gains in forecast accuracy are attributed to the forecast combination's success in attenuating model and parameter uncertainty. In fact, individual forecasting models are likely to be misspecified and this complicates identifying a best performing model (Elliott and Timmermann, 2016). In the context of predicting oil prices, Baumeister and Kilian (2015) show that combining oil price forecasts generates predictive gains relative to the no-change (i.e. random walk) forecast.

We test for forecast encompassing using a regression in differences so as to avoid spurious regression problems:

$$S_t - \hat{S}_t^{RW} = \beta(\hat{S}_t - \hat{S}_t^{RW}) + \varepsilon_t, \qquad (20)$$

where $\hat{S}_t$ is the forecast from one of the competing models and S_t^{RW} is the random walk forecast.

[35] We employ a size of 5% given the large sample size (for the in-sample and out-of-sample periods).

The null of forecast encompassing is given by $H_0 : \beta = 0$ versus the one-sided alternative $H_1 : \beta > 0$. For forecast combination to achieve statistical gains, the information in one forecast should not be encompassed (or subsumed) by the information contained in another forecast (Diebold, 2014).

Panel B of Table 4 provides the results from estimating the forecast encompassing regression in equation (19). The results show that the null that the random walk forecast encompasses the competing forecast cannot be rejected for any of the models except the ECM. In light of these results, we do not resort to forecast averaging and proceed to the trading exercise used to assess the economic significance of our results.

5. Direction Change Forecasting and Trading Strategies: Assessing Economic Significance

Since the contribution of Leitch and Tanner (1991), researchers recognize that statistical forecast accuracy does not invariably translate into economic profitability. In fact, Leitch and Tanner (1991) show that the correlation between statistical accuracy measures and profitability need not be positive or significant. Leitch and Tanner (1991) also note that forecasts which exhibit high directional accuracy are more likely to generate profits. Satchell and Timmermann (1995) offer arguments that are in line with those of Leitch and Tanner (1991). In this section, we assess the directional accuracy of our competing forecasts and design a trading strategy to examine the economic significance of our results. To do so, we employ crude oil price changes, Δs_t, as well as directional changes in the price of oil from one the competing forecasts/models $\Delta \hat{s}_t^i$.

5.1. Directional Forecast Accuracy

We begin by assessing the directional forecast accuracy of our models. To this end, we use the percent correct sign predictions given by:

$$\%.correct.sign = \frac{1}{T_f} \sum_{t=1}^{T_f} z_t, \qquad (21)$$

where $z_t = 1$ if $\Delta s_t . \Delta \hat{s}_{t+1} > 0$ and 0 otherwise. The % correct sign predictions are computed for our out-of-sample forecast evaluation period $t = 1, \ldots T_f$.

5.2. Trading Strategy and Transaction Costs

Trading signals are generated based on directional forecasts of logarithmic changes in predicted crude oil prices, $\Delta\hat{s}_t^i$. We adopt two simple trading strategies. Under the first, the investor goes long the first crude oil futures contract when the predicted crude oil price changes are positive. When the predicted oil price change is negative, the investor does not trade and earns the risk-free rate. We assume that the investor trades only directionally so as to minimize transaction costs. We refer to this strategy as the *long-only* strategy. Under the second, the investor goes short the front crude oil futures contract if the predicted return is negative and stays outside of the market, earning the risk-free rate, when the predicted crude oil price changes are negative. We refer to this strategy as *short-only* strategy.[36] The latter two strategies incur lower transaction costs than a *long-short* strategy since they require less frequent trading.

As a result of the high liquidity in the front crude oil futures contract, transaction costs are small. The unvaivalabity of bid and ask price data prevents obtaining estimates of transaction costs directly. To circumvent that, we rely on estimates of transaction costs reported in the existing literature.

Bessembinder, Carrion, Tuttle and Venkataraman (2014) report a mean effective (quoted) spread of 1.96 (1.13) basis points (or 0.0196% for the effective spread and 0.0113% for the quoted spread) for the front month (i.e. first or nearest) crude oil futures contract. We employ the quoted spread of Bessembinder, Carrion, Tuttle and Venkataraman (2014) of 0.0113% as our cost for a round trip and use the quoted half spread for a single trade (Bessembinder and Venkataraman, 2010). When a model predicts consecutive positive (negative) price changes, we assume that the long (short) position is maintained so as to minimize transaction costs.

5.3. Buy-and-Hold Strategy

This strategy assumes that the investor opens a long position in the first crude oil futures contract and closes the position at the end of the sample period. As noted before, we assume that the investor adopts a rolling over strategy under which she closes the open position in the nearest contract on

[36] Data on the daily risk-free rate are obtained from Kenneth French's data library. As a result of the Federal Reserve unconventional monetary policy actions (i.e. Large scale asset purchases or quantitative easing) the risk-free rate is very close to zero in our out-of-sample forecasting period.

the last day prior to the expiration month and opens a long position in the next-to-nearest contract. We further assume, for simplicity, that the investor does not incur any cost for rolling over from the front month to the second month futures contracts.

Our assumptions imply that the buy-and-hold strategy does not incur transaction costs. This, in turn, makes the buy-and-hold strategy a more stringent benchmark against which to compare the profitability of our econometric forecasts.[37] As a robustness check, we compute the returns to the buy-and-hold strategy using cash (spot) prices and our results are similar. However, we prefer using the first futures contract, as noted earlier, given that trading crude oil futures is simpler than transacting in the cash market.

5.4. Moving Average Convergence Divergence (MACD)

The moving average convergence divergence (MACD) is a popular technical analysis tool employed by traders. The performance of various variants of the MACD or momentum indicators has been thoroughly examined in trading commodity futures (Hurst, Ooi and Pedersen, 2013) and equities (Dbouk, Jamali and Soufani, 2014).

In a typical application, the use of the MACD requires the construction of three moving averages of crude oil price changes with different window sizes (i.e. number of days): a long, medium and a short moving average. We employ moving average window lengths of nine, twelve and twenty six days for the short, medium and long moving averages, respectively, given the popularity of this choice among traders (Dbouk, Jamali and Soufani, 2014).

After creating the three moving averages, we construct the MACD line as the difference between the long and short moving averages. Trading signals are generated as follows: the investor longs (shorts) the front crude oil futures when the MACD line crosses the intermediate moving average from below (above). As discussed before, we appropriately account for transaction costs every time a trade is executed. Akin to the trading strategies built from econometric forecasts, we consider long and short only strategies. Under the *long-only* (*short-only*) strategy, the investor acts upon a buy (sell) signal and stays outside the market (thereby earning the risk-free rate) otherwise.

Recent research suggests that the MACD strategy is successful in capturing time series momentum (Levine and Pedersen, 2016) and that time series momentum strategies can be

[37] Said differently, our assumptions are advantageous to a passive investor.

profitably used to trade commodity futures (Moskowitz, Ooi and Pedersen, 2012; Hurst, Ooi, Pedersen, 2013). These latter observations motivate our use of the MACD and suggest that the MACD can generate profits for futures traders.

5.5. Trading Profits and Directional Forecast Accuracy Results

Table 5 provides the percent correct sign predictions of our forecasts as well as the risk-adjusted returns (i.e. Sharpe ratios) of our trading strategies.

[Insert Table 5 here]

All of our linear and non-linear models succeed in outperforming the random walk in terms of directional forecast accuracy.

We commence our trading exercise by evaluating the realized returns of the different forecasts on a risk-adjusted basis. Following Dbouk, Jamali and Kryzanowski (2016), we compute the Sharpe ratio as:

$$\frac{E(\tilde{r}_{it})}{\sigma(\tilde{r}_{it})}, \tag{22}$$

where $\tilde{r}_{it}$ denotes the realized return from trading based on forecast i and $\sigma(\tilde{r}_{it})$ is the standard deviation of the realized return.

The results, also reported in Table 5, suggest that, with the exception of the MACD, all the Sharpe ratios for the *long-only* strategy are negative. These negative Sharpe ratios reflect the drop in crude oil prices which started towards the end of 2014. The buy-and-hold strategy, for example, generates a Sharpe ratio of -0.20. Only the MACD generates a larger risk-adjusted return than the buy-and-hold strategy. Our results also indicate that the Sharpe ratios for the *short-only* strategy are, with the exception of the MACD, negative. However, trading based on forecasts from the LINEAR 2, ARMA, ECM, ARDL, BANN, GANN and FANN provides a Sharpe ratio that is equal to or larger than the buy-and-hold strategy. The Sharpe ratios of the *short-only* strategy are consistently higher than those of the *long-only* strategy. This is expected given that the *short-only* strategy is designed to exploit the downward trend in oil prices.

The dollar payoffs, also reported in Table 5, provide similar conclusions as the Sharpe ratios. More specifically, only the terminal payoff of an initial \$100 investment accrues to \$115.66 for the MACD under a *long-only* strategy while the dollar payoffs of the remaining forecasts are lower than \$100. Again, the dollar payoffs are larger for the *short-only* strategy than the *long-only* strategy and trading based only on the MACD provides a terminal payoff greater than $100. Interestingly, while the buy-and-hold strategy yields dollar payoff that are larger than those of some of the linear models under a *long-only* strategy, all the models generate a higher dollar payoff than the buy-and-hold strategy under a *short-only* strategy. This implies that the models that we consider are better able to predict negative crude oil price changes than the random walk. This superior ability in predicting negative price changes translates into the higher profitability of the *short-only* strategy for the model forecasts.[38]

6. Concluding Remarks

This paper thoroughly examines the ability of linear and nonlinear models to predict the daily price of oil. We generate rolling one-step-ahead out-of-sample from linear and nonlinear models for the period January 4, 2010 to July 21, 2017. Two of the models that we employ are novel to the literature: the first is the demand model which exploits the informational content of copper prices as a proxy of global economic activity. The second, which we refer to as the financialization model, exploits the increased correlation between oil and equity prices as well as the predictive power of exchange rate changes. The nonlinear model that we consider is an artificial neural network. More specifically, we consider a neural network with bootstrap aggregation, a neural network trained using the genetic algorithm as well as a neural network combined with fuzzy logic.

After assessing the statistical accuracy of the competing forecasts, we examine the economic significance of the forecasts using a trading exercise. More specifically, we consider *long-only* and *short-only* trading exercises according to which an investor trades based on the sign of the predicted oil price change from one of the competing models. The profitability of the trading strategies is compared to that of the buy-and-hold strategy as well as to trading based on the moving average convergence divergence, a technical indicator.

We find that some of the linear models outperform the random walk in terms of out-of-sample statistical forecast accuracy. Our findings also suggest that while the buy-and-hold strategy dominates some of the models in terms of dollar payoffs and risk-adjusted returns under a *long-only* strategy, all the models that we consider generate higher dollar payoffs than the buy-and-hold strategy under the *short-only* strategy. We also find that the profits from the *short-only* strategy

[38] Pesaran and Timmermann (1994) note that the ability of a model to accurately predict negative changes is an important determinant of its profitability.

are consistently larger than those of the *long-only* strategy. This likely reflects the superior ability of the linear and nonlinear models, vis-à-vis the random walk, in predicting negative oil price changes, which became more frequent with the decline in the price of oil at the end of 2014. Trading based on the MACD, a technical indicator which adequately captures trends, generate the largest payoff among all the competing models.

References

Alquist, R., Kilian, L., & Vigfusson, R. (2013). Forecasting the price of oil. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Vol. 2 (pp. 427–507). Amsterdam: North-Holland.

Bakshi, G., Panayotov, G., & Skoulakis, G. (2011). The Baltic dry index as a predictor of global stock returns, commodity returns, and global economic activity. Working paper, University of Maryland.

Baumeister, C., Guérin, P., & Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. *International Journal of Forecasting*, *31*, 238-252.

Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business and Economic Statistics*, *30*, 326-336.

Baumeister, C., & Kilian, L. (2015). Forecasting the real price of oil in a changing world: a forecast combination approach. *Journal of Business and Economic Statistics, 33*, 338-351.

Baumeister, C., Kilian, L., & Zhou, X. (2016). Are product spreads useful for forecasting oil prices? An empirical evaluation of the Verleger hypothesis, *Macroeconomic Dynamics*, forthcoming.

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics, 183*, 181-192.

Bessembinder, H. (1992). Systematic risk, hedging pressure, and risk premiums in futures markets. *Review of Financial Studies*, *5*, 637-667.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., & Smoller, M. M. (1996). Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis. *Journal of Derivatives*, *4*, 45-58.

Bessembinder, H., & Venkataraman, K. (2010). Bid-ask spreads. *Encyclopedia of Quantitative Finance*, 1–6.

Bera, A. K., & Jarque, C. M. (1981). An efficient large-sample test for normality of observations and regression residuals. Australian National University, Faculty of Economics and Research School of Social Sciences.

Bernanke, B. S. (2016). The relationship between stocks and oil prices. https://www.brookings.edu/2016/02/19/the-relationship-between-stocks-and-oil-prices/

Bessembinder, H., Carrion, A., Tuttle, L., & Venkataraman, K. (2014). Predatory or sunshine trading? evidence from crude oil ETF rolls. Working paper, University of Utah.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *Review of Financial Studies, 22*, 4463-4492.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123-140.

Brooks, C., Rew, A. G., & Ritson, S. (2001). A trading strategy based on the lead–lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting*, *17*, 31-44.

Burghardt, G. (2008). Volume surges again: Global futures and options trading rises 28% in 2007, *Futures Industry Magazine*, March/April, 15-26.

Buckley, J. J., & Hayashi, Y. (1994). Fuzzy neural networks: A survey. *Fuzzy Sets and Systems*, *66*, 1-13.

Buyuksahin, B., & Robe, M. A. (2011). Does paper oil matter? Energy markets' financialization and equity-commodity co-movements. *Energy Markets' Financialization and Equity-Commodity Co-Movements (July 28, 2011).*

Caldara, D., Cavallo, M., & Iacoviello, M. (2016). Oil price elasticities and oil price fluctuations. International Finance Discussion Paper Number 1173, Board of Governors of the Federal Reserve System.

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

Chen, S-S. (2009). Oil price pass-through into inflation. *Energy Economics, 31*, 126-33.

Chen, Y. C., Rogoff, K., & Rossi, B. (2010). Can exchange rates forecast commodity prices? *Quarterly Journal of Economics*, *125*, 1145-1194.

Cheng, I-H, & Xiong, W., (2014). Financialization of commodity markets, A*nnual Review of Financial Economics, 6*, 419-441.

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

Coppola, A. (2008). Forecasting oil price movements: Exploiting the information in the futures market. *Journal of Futures Markets*, *28*, 34-56.

Dbouk, W., Jamali, I., & Kryzanowski, L. (2016). Forecasting the LIBOR-Federal Funds Rate Spread During and After the Financial Crisis. *Journal of Futures Market*, *36*, 345-374.

Dbouk, W., Jamali, I., & Soufani, K. (2014). The effectiveness of technical trading for Arab stocks. *Emerging Markets Finance and Trade*, *50*, 5-25.

De Roon, F. A., Nijman, T. E., & Veld, C. (2000). Hedging pressure effects in futures markets. *Journal of Finance 55,* 1437-1456.

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

Diebold, F. X. (2014). *Forecasting in Economics, Business, Finance and Beyond*. University of Pennsylvania.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*, 813-836.

Elliott, G., & Timmermann, A., (2006). *Economic Forecasting*, Princeton University Press.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, *55,* 251-276.

Fama, E. F., & French, K. R. (1987). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. *Journal of Business*, *60,* 55-73.

Fama, E. F., & French, K. R. (1988). Business cycles and the behavior of metals prices. *Journal of Finance*, *43*, 1075-1093.

Fattouh, B., Kilian, L. & Mahadeva, L. L. (2013). The role of speculation in oil markets: What have we learned so far? *Energy Journal*, *34*, 7-33.

Franses, P. H., & Van Dijk, D. (2002). *Nonlinear Time-Series Models in Empirical Finance*, Cambridge University Press.

Franses, P. H., & Van Griensven, K. (1998). Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics & Econometrics*, *2*, 109-114.

Franses, P. H., & Van Homelen, P. (1998). On forecasting exchange rates using neural networks. *Applied Financial Economics*, *8,* 589-596.

Giacomini, R., & White, H., (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545-1578.

Gorton, G., & Rouwenhorst, K. G. (2006). Facts and fantasies about commodity futures. *Financial Analysts Journal*, *62,* 47-68.

Gospodinov, N. (2016). The role of commodity prices in forecasting US core inflation. *Federal Reserve Bank of Atlanta Working Paper No. 2016-5*.

Gospodinov, N., & Jamali, I. (2011). Risk premiums and predictive ability of BAX futures. *Journal of Futures Markets, 31*, 534-561.

Gospodinov, N., & Wei, B. (2016). Forecasts of inflation and interest rates in no-arbitrage affine models, Working Paper 2016-3, Federal Reserve Bank of Atlanta.

Gradojevic, N., Gençay, R., & Kukolj, D. (2009). Option pricing with modular neural networks. *IEEE transactions on neural networks*, *20*, 626-637.

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

Hamilton, J. D. (1985). Historical causes of postwar oil shocks and recessions. *The Energy Journal*, *6*, 97-116.

Hamilton, J. D. (2003). What is an oil shock?. *Journal of Econometrics, 113,* 363-398.

Hamilton, J. D. (2009). Causes and Consequences of the Oil Shock of 2007-08, *Brookings Papers on Economic Activity, 1*, 215-259.

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

Hamilton, J. D. (2013). Historical Oil Shocks. In. R. E. Parker & R. Whaples (Eds.) *Handbook of Major Events in Economic History* (pp. 239-265), New York: Routledge Taylor and Francis Group.

Hamilton, J.D. (2015). Demand factors in the collapse of oil prices. http://econbrowser.com/archives/2015/01/demand-factors-in-the-collapse-of-oil-prices accessed via www.Econbrowser.com.

Hamilton, J. D., & Herrera, A. M. (2004). Oil shocks and aggregate macroeconomic behavior: the role of monetary policy: a comment. *Journal of Money, Credit, and Banking*, *36*, 265-286.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281-291.

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2013). Demystifying managed futures. *Journal of Investment Management*, *11*, 42-58.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. *Journal of the American Statistical Association*, *103*, 511-522.

Irwin, S. H., & Sanders, D. R. (2012). Testing the Masters hypothesis in commodity futures markets. *Energy Economics*, *34*, 256-269.

Khwaja, A. S., Naeem, M., Anpalagan, A., Venetsanopoulos, A., & Venkatesh, B. (2015). Improved short-term load forecasting using bagged neural networks. *Electric Power Systems Research*, *125*, 109-115.

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

Kilian, L. (2010). Explaining fluctuations in gasoline prices: a joint model of the global crude oil market and the US retail gasoline market. *Energy Journal*, *31*, 87-104.

Kilian, L. (2014). Oil Price Shocks: Causes and Consequences, *Annual Review of Resource Economics*, *6*, 133-154.

Kilian, L., & Vigfusson, R. J. (2011a). Are the responses of the U.S. economy asymmetric in energy price increases and decreases? *Quantitative Economics*, *2*, 419-453.

Kilian, L., & Vigfusson, R. J. (2011b). Nonlinearities in the oil price-output, *Macroeconomic Dynamics*, *15*, 337-363.

Kilian, L., & Vigfusson, R. J. (2013). Do oil prices help forecast U.S. real GDP? The role of nonlinearities and asymmetries. *Journal of Business and Economic Statistics*, *31*, 78-93.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, *10*, 169-181.

Kryzanowski, L., Galler, M., & Wright, D. W. (1993). Using artificial neural networks to pick stocks. *Financial Analysts Journal*, *49*, 21-27.

Kuan, C-H, & White, H., (2012). Artificial neural networks: An econometric perspective. *Econometric Reviews*, *13*, 1-91.

Labys, W., Achouch, A. & Terraza, M. (1999). Metal prices and the business cycle. *Resources Policy*, *25*, 229-238.

Leitch, G., & Tanner, J. E. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review*, *81*, 580-590.

Levine, A., & Pedersen, L. H. (2016). Which trend is your friend?. *Financial Analysts Journal*, *72*, 51-66.

Maasoumi, E., Khotanzad, A., & Abaye, A., 1994. Approximation and radial basis networks for some macroeconomic series: A first report. Econometric Reviews, *13*, 105–122.

McNelis, P. D. (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Academic Press.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* (pp. 3-46). National Bureau of Economic Research.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*, 228-250.

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, *86*, 373-378.

Newey, W. K., & West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703-708.

Novák, V., Perfilieva, I., & Mockor, J. (2012). *Mathematical Principles of Fuzzy Logic*, Springer Science & Business Media.

Pesaran, M. H., & Timmermann, A. (1994). Forecasting stock returns An examination of stock market trading in the presence of transaction costs. *Journal of Forecasting*, *13*, 335-367.

Pindyck, R. S., & Rotemberg, J. J. (1990). The excess co-movement of commodity prices. *Economic Journal*, *100*, 1173-1189.

Rapach, D. E., & Strauss, J. K. (2010). Bagging or combining (or both)? An analysis based on forecasting US employment growth. *Econometric Reviews*, *29*, 511-533.

Refenes, A. N., Zapranis, A., & Francis, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. *Neural networks*, *7*, 375-388.

Satchell, S., & Timmermann, A. (1995). An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting*, *14*, 477-497.

Schinas, O., Grau, C., & Johns, M. (Eds.). (2015). *HSBA Handbook on Ship Finance*. Springer.

Skiadopoulos, G. (2013). Advances in the commodity futures literature: A review. *Journal of Derivatives*, *20*, 85-96.

Stoll, H. R., & Whaley, R. E. (2010). Commodity index investing and commodity futures prices. *Journal of Applied Finance*, *20*, 1-40.

Singleton, K. J. (2013). Investor flows and the 2008 boom/bust in oil prices. *Management Science*, *60*, 300-318.

Swanson, N. R. & White, H., (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks, *Journal of Business and Economic Statistics*, *13*, 265-275.

Swanson, N. R., & White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics*, *79*, 540-550.

Szafranek, K. (2017). *Bagged artificial neural networks in forecasting inflation: An extensive comparison with current modelling frameworks*. Narodowy Bank Polski, Economic Research Department.

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, *68,* 54-74.

Teräsvirta, T., Van Dijk, D., & Medeiros, M. C. (2005). Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, *21*, 755-774.

Theil, H. (1966). Applied econometric forecasting. *Amsterdam, NE: North-Holland*.

Timmermann, A. (2006). Forecast combinations, in G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting,* 1, (pp. 135-196) Amsterdam, The Netherlands: Elsevier.

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

Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management, 26*, 12-17.

White, H., 1988. Economic prediction using neural networks: The case of IBM daily stock returns. Proceedings of the IEEE International Conference on Neural Networks, 2, 451–458.

Yu, L., Wang, S., & Lai, K. K. (2010). *Foreign Exchange Rate Forecasting with Artificial Neural Networks*, Springer Science & Business Media.

Zhang, X. (1994). Time series analysis and prediction by neural networks. *Optimization Methods and Software, 4*, 151-170.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting, 14*, 35-62

Zhou, Z. H., Wu, J., & Tang, W. (2002). Ensembling neural networks: many could be better than all. *Artificial intelligence*, *137*, 239-263.

TABLE 1 Unit Root Tests and Descriptive Statistics for Crude Oil Prices and Price Changes

Panel A: Unit Root Tests for Crude Oil Prices						
	ADF			ADF-GLS		
Full Sample	-2.00			-1.98		
In-Sample Period	-2.31			-1.77		
Out-of-Sample Period	-2.09			-1.42		
Panel B: Unit Root Tests for Log Crude Oil Prices						
	ADF			ADF-GLS		
Full Sample	-2.34			-2.11		
In-Sample Period	-2.77			-1.88		
Out-of-Sample Period	-2.11			-1.53		
Panel C: Summary Statistics for Crude Oil Returns						
	Mean	Median	Std. Dev.	AC(1)	Skewness	Kurtosis
Full Sample	0.01	0.00	2.38	-0.01	-0.73	18.58
In-Sample Period	0.02	0.00	2.49	-0.00	-0.91	20.32
Out-of-Sample Period	-0.02	0.00	2.06	-0.05	0.09	6.09

Notes: Panel A provides the unit root tests of the oil price series. ADF refers to the Augmented Dickey Fuller Test of Dickey and Fuller (1979). The ADF-GLS test refers to the ADF test with GLS detrending of Elliott, Rothenberg and Stock (1996). Panel B provides the summary statistics of crude oil returns. AC(1) refers to the first-order autocorrelation and Std. Dev. refers to the standard deviation.

TABLE 2 Summary Statistics and Cross-Correlations of Predictors

Panel A: Summary Statistics for Predictors

	Mean	Median	Std. Dev.	AC(1)	Skewness	Kurtosis
ΔPCOM	0.007	0.000	0.408	0.064	-0.459	9.045
ΔEXRATE	-0.003	0.003	0.435	-0.001	-0.239	6.475
ΔYIELD	-0.001	0.000	0.057	0.027	0.034	5.260
ΔCOPPER	0.013	0.000	1.627	-0.062	-0.190	7.321
ΔVIX	-0.000	-0.010	1.492	-0.091	0.685	21.871
ΔBDI	-0.005	0.027	0.410	0.998	-1.744	11.335

Panel B: Cross-Correlations of Predictors

	ΔPCOM	ΔEXRATE	ΔYIELD	ΔCOPPER	ΔVIX	ΔBDI
ΔPCOM	1.000	-	-	-	-	-
ΔEXRATE	-0.223	1.000	-	-	-	-
ΔYIELD	0.099	0.045	1.000	-	-	-
ΔCOPPER	0.363	-0.245	0.134	1.000	-	-
ΔVIX	-0.118	0.047	-0.169	-0.194	1.000	-
ΔBDI	0.063	-0.006	0.029	0.041	-0.006	1.000

Notes: The table provides the summary statistics and cross-correlations of crude oil price changes. *ΔPCOM* refers to the logarithmic change (in percent) in the CRB spot commodity price index. *ΔEXRATE* is the logarithmic change (in percent) in the U.S. dollar trade-weighted exchange rate. *ΔYIELD* is the change in the yield on the U.S. ten year Treasury note. *ΔCOPPER* is the logarithmic change (in percent) in the prices of the nearest copper futures contract. VIX is the level of the CBOE option-implied volatility index. *ΔBDI* is the three-month logarithmic change (in percent) in the Baltic Dry Index. AC(1) denotes the first-order autocorrelation and Std. Dev. denotes the standard deviation.

TABLE 3 Statistical Forecast Accuracy and Testing Differences in Forecast Accuracy

Panel A: Statistical Forecast Accuracy Measures

	MAE	RMSE	MAPE	Theil U
Random Walk	1.0154 (3)	1.3775 (5)	0.0146 (3)	1.0000
Linear 1	1.0308 (7)	1.3741 (3)	0.0148 (7)	**0.9975**
Linear 2	1.0583 (9)	1.4156 (9)	0.0151 (9)	1.0276
Demand	1.0045 (2)	1.3445 (2)	0.0145 (2)	**0.9760**
ARMA	1.0214 (5)	1.3812 (7)	0.0147 (5)	1.0026
ECM	1.0175 (4)	1.3745 (4)	0.0147 (4)	**0.9961**
ARDL	1.0405 (8)	1.4000 (8)	0.0150 (8)	1.0163
Financialization	0.9978 (1)	1.3401 (1)	0.0144 (1)	**0.9728**
BANN	6.9012 (11)	13.8974 (11)	0.0717 (11)	10.0895
GAANN	2.4081 (10)	3.8655 (10)	0.0281 (10)	2.8063
FANN	1.0222 (6)	1.3796 (6)	0.0147 (6)	1.0000

Panel B: Tests of Statistical Differences in Forecast Accuracy

	OoS	GW	MDM	MSPE-adjusted
Random Walk	-	-	-	-
Linear 1	**0.49**	**20.00**$^{(-)}$	-3.64	0.54
Linear 2	-5.60	**17.61**$^{(-)}$	-4.95	-0.36
Demand	**4.73**	0.10	-0.43	1.19
ARMA	-0.53	1.66	-1.23	0.02
ECM	**0.43**	1.54	1.23	2.57***
ARDL	-3.29	**17.56**$^{(-)}$	-1.15	-3.58
Financialization	**5.35**	0.35	-0.46	0.74
BANN	-10079.87	**373.34**$^{(-)}$	-4.17	1.15
GAANN	-687.58	**337.17**$^{(-)}$	-4.63	1.21
FANN	-0.32	2.54	-0.52	1.30*

Notes: The table provides the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error, Theil *U* and Out-of-Sample (OoS) R^2 statistic for each of the competing forecasts. The Theil *U* statistics that are less than one are in bold. Positive OoS statistics are in bold. The column *GW* reports the conditional version of the Giacomini and White (2006) statistic under quadratic loss. The *GW* test compares the predictive accuracy of a model to that of the random walk benchmark. A plus (minus) sign for *GW* indicates that the model outperforms (underperforms) the benchmark statistically significantly at the 10% level. The number in parentheses is the rank of the model/forecast according to the criterion chosen. *MDM* refers to the Harvey, Leybourne and Newbold (1997) modified version of the Diebold and Mariano (1995) test. *MSPE-adjusted* is the Clark and West (2007) statistic, which modified the Diebold and Mariano (1995) test to allow for comparing nested models. *, ** and *** denote, respectively, statistical significance at the 10%, 5% and 1% levels.

TABLE 4 Forecast Optimality and Encompassing Tests

Panel A: Tests of Forecast Optimality

	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	*p*-value
Random Walk	0.082	0.094	0.998	0.001	0.52
Linear 1	0.153	0.101	0.997	0.001	0.22
Linear 2	0.104	0.090	0.998	0.001	0.33
Demand	0.173	0.096	0.997	0.001	0.06
ARMA	0.086	0.098	0.998	0.001	0.30
ECM	-0.060	0.097	1.000	0.001	0.80
ARDL	0.016	0.106	0.995	0.001	0.80
Financialization	0.154	0.095	0.997	0.001	0.09
BANN	4.685	1.525	1.019	0.019	0.00
GAANN	-3.525	0.406	1.067	0.007	0.00
FANN	-0.114	0.103	1.001	0.001	0.53

Panel B: Tests of Forecast Encompassing

	$\hat{\beta}$	$se(\hat{\beta})$	*p*-value
Random Walk	-	-	-
Linear 1	0.057	0.104	0.58
Linear 2	-0.048	0.113	0.66
Demand	0.397	0.271	0.14
ARMA	0.012	0.402	0.97
ECM	0.858	0.317	0.00
ARDL	-0.187	0.161	0.24
Financialization	0.305	0.417	0.46
BANN	0.002	0.002	0.20
GAANN	0.010	0.008	0.18
FANN	0.369	0.254	0.46

Notes: The table provides the results from the forecast optimality and forecast encompassing tests in equations (18) and (19). The *p*-value in Panel A refers to the *F*-statistic used to test the null $H_0 : \beta_0 = 0, \beta_1 = 1.$ The *p*-value in Panel B refers to the *t*-statistic used to test the null $H_0 : \beta = 0.$ Heteroskedasticity and autocorrelation consistent standard errors of Newey and West (1987) standard errors (with automatic lag selection) are reported next to each coefficient.

TABLE 5 Directional Forecast Accuracy and Trading Strategy Profitability

		Payoff to $100		Sharpe Ratio	
	% Correct Sign	Long Only	Short Only	Long Only	Short Only
Random Walk	44.74	-	-	-	-
Linear 1	47.47	$45.29	$57.13	-0.35	-0.23
Linear 2	46.93	$39.83	$71.76	-0.40	-0.06
Demand	46.52	$51.95	$51.79	-0.28	-0.29
ARMA	46.22	$31.73	$56.51	-0.54	-0.19
ECM	47.13	$41.26	$70.93	-0.42	-0.07
ARDL	46.52	$32.36	$56.59	-0.53	-0.19
Financialization	46.45	$44.57	$54.51	-0.39	-0.25
BANN	48.40	$41.27	$71.11	-0.39	-0.06
GAANN	47.13	$39.55	$68.38	-0.41	-0.08
FANN	46.97	$40.60	$70.04	-0.40	-0.06
MACD	-	$115.66	$144.94	0.15	0.31
Buy-and-Hold	-	$38.77	$38.77	-0.20	-0.20

Notes: The table provides the percent correct sign predictions, Sharpe ratios and terminal values of a $100 investment to *long-only* and *short-only* strategies based on the directional forecasts from the competing models. The trading strategy profits are compared to the profitability of trading based on a Moving Average Convergence Divergence (MACD) and Buy-and-Hold strategies.