PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE

ESCUELA DE INGENIERIA

HOW GOOD ARE ANALYST FORECASTS OF OIL PRICES?

CONSUELO A. VALENCIA

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Advisor:

GONZALO CORTAZAR

Santiago de Chile, (May, 2020)

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

HOW GOOD ARE ANALYST FORECASTS OF OIL PRICES?

CONSUELO A. VALENCIA

Members of the Committee:

GONZALO CORTAZAR

TOMAS REYES

HECTOR ORTEGA

JULIO VERGARA

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Santiago de Chile, May 2020

To my family and schoolfellows

ACKNOWLEDGMENTS

I would like to express my gratitude to the advisor of this work, professor Gonzalo Cortazar for his invaluable knowledge transmitted. Thanks also to Hector Ortega, for all the patience and support during this process.

Likewise, I thank the members of the Committee Julio Vergara, and Tomas Reyes for their time during the evaluation process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Even though there is a wide consensus that having good oil price forecasts is very valuable for many agents in the economy, results have not been fully satisfactory and there is an ongoing effort to improve their accuracy. Research has explored many different modeling approaches including time series, regressions, and artificial intelligence, among others. Also many different sources of input data have been used like spot and futures prices, product spreads, and micro and macro variables.

This paper explores how useful are analyst´s expected price data for forecasting when appropriate measures are taken to account for sparse nature and high volatility. It proposes a multifactor stochastic pricing model, with time-varying risk premiums calibrated with filtered futures and analyst´s forecasts using a Kalman Filter.

The forecasting model is applied to ten years of oil prices and analyst forecasts, from NYMEX and Bloomberg, respectively. Results are very encouraging showing that the model's forecast performs much better than the no-change forecast, commonly used as a benchmark and better than many of the forecasting results from the literature. We conclude that analyst forecasts is a valuable source of input data that should be considered in future forecasting models.

Keywords: Forecasting, Oil prices; Futures; Expected Prices; Pricing Models.

RESUMEN

Si bien existe un amplio consenso en que tener buenos pronósticos del precio del petróleo es muy valioso para muchos agentes de la economía, los resultados no han sido completamente satisfactorios y existe un permanente esfuerzo por mejorar su precisión. La investigación ha explorado muchos modelos diferentes, incluyendo series de tiempo, regresiones e inteligencia artificial, entre otros. También se han usado bastantes fuentes asociadas a datos de entrada, como precios *spot* y futuros, diferenciales de producto, y tanto micro como macro variables.

Este trabajo explora cuán útiles son los datos de precios esperados de los analistas para predecir cuando se toma las medidas apropiadas, que permiten considerar la naturaleza dispersa y volátil de estos datos. Se propone un modelo de precios estocástico multifactorial, con primas de riesgo variables en el tiempo calibradas con futuros filtrados y pronósticos de analistas utilizando un filtro de Kalman.

El modelo de predicción de precios evalua diez años de precios de petróleo y pronósticos de analistas, de NYMEX y Bloomberg, respectivamente. Los resultados son bastante alentadores, muestran que el pronóstico del modelo planteado se comporta mucho mejor que el asociado al comunmente usado como *benchmark* (e.g. *no-change*) y, mejor que varios modelos de predicción de precios de la literatura. Concluimos que los pronósticos de precios de los analistias son una fuente valiosa de datos que deben considerarse en futuros modelos de predicción de precios.

Palabras Claves: Predicción, Precios Petróleo; Futuros; Expectativas de precios; Modelos, Predicción Precios.

I. ARTICLE BACKGROUND

There is a wide consensus that having good oil price forecasts is very valuable for agents in economies that import or export this resource. Also, it is a main factor for macroeconomic projections by central banks and for private sector growth estimations (Alquist, Kilian, & Vigfusson, 2013). The price of oil and its derivatives impact the consumption behavior of energy-intensive durable goods (Busse, Knittel, & Zettelmeyer, 2009), and can challenge monetary and fiscal policies (Baffes, Kose, Ohnsorge, & Stocker, 2015).

Commodity prices are variable and not trivial to predict. Specifically, oil price is subject to unexpected fluctuations since its influenced by multiple factors like weather, stock levels, GDP growth, political aspects and people's psychological expectations (Yu, Wang, & Lai, 2008). Given their importance, research has explored many different modeling approaches including time series (e.g. vector autoregression, VAR techniques), regressions, and Artificial Intelligence (AI), among others. Also many different sources of data have been used like spot and futures prices, product spreads, and micro and macro variables. Most forecasting models use as their benchmark a random-walk or no-change forecast to evaluate their performance.

This work explores how useful are analyst´s expected price data for forecasting when appropriate measures are taken to account for sparse nature and high volatility. It proposes a multifactor stochastic pricing model, with time-varying risk premiums calibrated with filtered futures and analyst´s forecasts using a Kalman Filter.

In order to compare the approach associated with the newly named stochastic model, we describe some of the techniques of the forecasting literature. One common approach is to propose a vector of autoregression (VAR) model like the one in Kilian and Murphy (2014) that considers shocks to the speculative demand of oil (using data on oil inventories), shocks to the flow of demand and supply. A reduced-form representation of this structural global oil market model with 12 autoregressive lags is used by Baumeister & Kilian (2012) and Baumeister, Kilian & Lee (2014), showing improved forecast accuracy in some pooled forecasts at very short term horizons.

Inspired by the fact that oil is not a homogenous commodity, another time series approach is proposed by Lanza, Manera and Giovannini (2005) that applies an error correction model (ECM) in order to anticipate the evolution of crude oil prices. They analyze the dynamic relationships among heavy crude oil and product prices, using an autoregressive-distributed lag specification that – after testing cointegration between the oil and product price series - admit an ECM formulation. Their results show that product prices are statistically relevant in explaining short - and long - run adjustment in petroleum markets.

A classical forecasting approach is to use regression techniques. Wang, Liu, Diao and Wu (2017) propose a two-step method to improve the predictability of real oil price by handling the problem of overfitting caused by redundant variables. Miao, Ramchander, Wang and Yang (2017) uses the Least Absolute Shrinkage and Selector Operator (LASSO) in their regression model to forecast oil prices as a method to choose the predictive variables.

Many forecasting models are now relying on some form of AI. One approach includes neural networks (Sun, Sun, Wang, & Wei, 2018; Cheng, Li, Wei, & Fan, 2019; Zhang, Zhang, & Zhang, 2015). Based on the "divide-and-conquer" principle, first, they decompose – in order to simplify the forecasting aim, they use independent sub-series – and then, ensemble to formulate a consensus forecasting on the original data. Yu et al. (2008) proposes an empirical mode decomposition (EMD) based on neutral networks ensemble learning paradigm for oil spot price forecasting.

Related with a machine learning approach, Zhao, Li and Yu (2017) uses AI to forecast oil prices. In summary, first they divide the data into two parts, training and test samples, and model the nonlinear relationships of oil price with many factors. Then one prediction for each trained set is done to finally take the average prediction value as the final forecast.

In a recent work Li, Zhu, & Wu (2019) implement hybrid methods based on variational mode decomposition (VMD) and artificial intelligence (AI) techniques for forecasting the trend component in monthly crude oil prices. It is important to note that in this work, influencing factors of long-term crude oil price variation are considered for the forecasting of the trend subcomponent.

Forecasting models in the literature diverge not only on the modeling approach but also on the data used. Having said that, we recognize many of them using past spot prices, some of them use futures prices and/or macroeconomic and financial series, including financial indicators such as indexes of industry, stock and future market, gold price, interest rates and dollar indices. Other forecasting models use product spreads (Baumeister, Kilian, & Zhou, 2018) such as gasoline and heating oil spot spreads. This is based on the idea that gasoline prices and crude oil prices move together in the long run (Lanza et al., 2005). They showed that not all product spread models are useful for out-of-sample forecasting, but the most accurate single spread forecasting model is based on the gasoline spot spread alone.

A less common input to the forecasting models is to use analyst´s expectations. There has been some research on the usefulness of expectations surveys. Alquist et al. (2013) evaluates the performance of survey forecasts of the nominal price of oil and shows that there is no compelling evidence that survey forecasts outperform the no-change forecast. Gay, Simkins, & Turac (2009) analyzes how the market incorporates Bloomberg analysts' forecasts in the price of natural gas. Stark (2010) focuses on how accurate are the Survey of Professional Forecasters (SPF) predictions of inflation, unemployment, concluding that in general the SPF outperform the benchmark projections of univariate autoregressive time series at short horizons. Pierdzioch, Rülke, & Stadtmann (2010) find strong evidence of anti-herding among oil-price forecasters. Cortazar, Kovacevic & Schwartz (2015) validates model predictions using expectation surveys. Cortazar, Millard, Ortega, & Schwartz (2019) use futures and expectation surveys to estimate a constant term structure of oil risk premiums while Cortazar, Liedtke, Ortega, & Schwartz (2018) uses the same information to estimate a time-varying term structure of oil premiums and what are the micro and macro variables that explain their variation.

As said before, in this paper we explore the usefulness of including a survey of analysts´ forecasts as input in an oil price forecasting model. We address the high volatility of expectation forecasts by including also futures price data, by using a Kalman filter to allow for observational error and by using the three factor stochastic model that allows for a time-varying risk premium term structure. This model was proposed by Cortazar et al. (2018) and we use ten years of crude oil prices from NYMEX and analyst forecasts from Bloomberg as data input for this model.

Performance of forecasting models are typically measured using the Relative MSPE and Directional Accuracy and compared with a no-change forecast, commonly used as a benchmark. Under the present formulation Relative MSPEs below 1 mean that the forecast is more accurate than the no-change forecast. Regarding Directional Accuracy, this metric is evaluated using the Success Ratio, which represent the average number of times that the model success in the direction of its prediction. Given that, any Success Ratio higher than 0.5 indicates an improvement over the no-change forecast.

In order to make a quick comparison with some results published from other models in the literature, we choose six alternative models to analyze their results. In terms of the Relative MSPE metrics, our model is better than all the other models for all horizons, except for the one and three month periods. Regarding the Directional Accuracy metric, results of our model are even better. Our model overcomes all others for all horizons except for the 3 month forecasts. It is important to the objective of this analysis was not to determine the best model, but rather to see if results are reasonably close to those of other models thus to determine if there is valuable information in the analysts´ expected prices.

Using the aforementioned metrics, our results are very encouraging showing that the model performs much better than the no-change model, and also better than many of the forecasting models from the literature. The Relative MSPE results are below 1 for every horizon in analysis, standing out the 18, 21 and 24 months horizons, with a ratio below 0.7 for this metric. The Relative MSPE for horizons up to one and two years are both below 0.9. Regarding the Directional Accuracy results, these are evaluated through the Success Ratio which value goes from 0.55 to 0.82 for the horizons in analysis, all being statistically significant. If we consider the horizons up to one and two years this ratio is above 0.59 for both of them.

Our main conclusion is that using analysts´ expected prices, even if they are noisy, is a useful data source that should be taken into account in the future by commodity price forecasters. Analyst forecasts is a valuable source of input data that should be considered in future forecasting models.

II. HOW GOOD ARE ANALYST FORECASTS OF OIL PRICES?

1. Introduction

There is a wide consensus that having good oil price forecasts is very valuable for agents in economies that import or export this resource. Also, it is a main factor for macroeconomic projections by central banks and for private sector growth estimations (Alquist, Kilian, & Vigfusson, 2013). The price of oil and its derivatives impact the consumption behavior of energy-intensive durable goods (Busse, Knittel, & Zettelmeyer, 2009), and can challenge monetary and fiscal policies (Baffes, Kose, Ohnsorge, & Stocker, 2015).

Oil prices are difficult to predict because they are very volatile and subject to unexpected fluctuations of many factors like weather, stock levels, GDP growth, political issues and personal expectations (Yu, Wang, & Lai, 2008). Given their importance, research has explored many different modeling approaches including time series (e.g. vector autoregression, VAR techniques), regressions, and Artificial Intelligence (AI), among others. Also many different sources of data have been used like spot and futures prices, product spreads, and micro and macro variables. Most forecasting models use as their benchmark a random-walk or no-change forecast to evaluate their performance.

This paper explores how useful are analyst´s expected price data for forecasting when appropriate measures are taken to account for their sparse nature and high volatility. It proposes a multifactor stochastic pricing model, with time-varying risk premiums calibrated with filtered futures and analyst´s forecasts using a Kalman Filter.

To put our proposal into perspective we describe some of the approaches of the forecasting literature. One common approach is to propose a VAR model like the one in Kilian and Murphy (2014) that considers shocks to the speculative demand of oil (using data on oil inventories) and shocks to the flow of supply. A reduced-form representation of this structural global oil market model with 12 autoregressive lags is used by Baumeister & Kilian (2012) and Baumeister, Kilian & Lee (2014), showing improved forecast accuracy in some pooled forecasts at short term horizons.

Another time series approach is proposed by Lanza, Manera and Giovannini (2005) that applies an error correction model (ECM) in order to anticipate the evolution of oil prices. They analyze the dynamic relationships among heavy crude oil and product prices, using an autoregressive-distributed lag specification that – after testing for cointegration between the oil and product price series - admit an ECM formulation.

A classical forecasting approach is to use regression techniques. Wang, Liu, Diao and Wu (2017) propose a two-step method to improve the predictability of real oil price by handling the problem of overfitting caused by redundant variables. Miao, Ramchander, Wang and Yang (2017) uses the Least Absolute Shrinkage and Selector Operator (LASSO) in their regression model to forecast oil prices as a method to choose the predictive variables.

Many forecasting models are now relying on some form of AI. One of the most common approaches includes neural networks (Sun, Sun, Wang, & Wei, 2018; Cheng, Li, Wei, & Fan, 2019; Zhang, Zhang, & Zhang, 2015). Based on the "divide-and-conquer" principle, first, they decompose – in order to simplify the forecasting aim, they use independent sub-series – and then, ensemble to formulate a consensus forecasting on the original data. Yu et al. (2008) proposes an empirical mode decomposition (EMD) based on neural networks ensemble learning paradigm, for oil spot price forecasting.

Zhao, Li and Yu (2017) uses a machine learning approach to AI to forecast oil price. In summary, first they divide the data into two parts, training and test samples, and model the nonlinear relationships of oil price with many factors. Then one prediction for each trained set is done to finally take the average prediction value as the final forecast.

In a recent work Li, Zhu, & Wu (2019) implement hybrid methods based on variational mode decomposition (VMD) and artificial intelligence (AI) techniques for forecasting the trend component in monthly crude oil prices.

Forecasting models in the literature diverge not only on the modeling approach but also on the data used. Many of the forecasting models use past spot prices. Some of them use futures prices and some macroeconomic and financial series, including financial indicators such as indexes of industry, stock and future market, gold price, interest rates and dollar indices. Some other forecasting models use product spreads (Baumeister, Kilian, & Zhou, 2018) such as gasoline and heating oil spot spreads. This is based on the idea that gasoline prices and crude oil prices move together in the long run (Lanza et al., 2005).

A less common input to the forecasting models is to use analyst´s expectations. There has been some research on the usefulness of expectations surveys. Alquist et al. (2013)

evaluates the performance of survey forecasts of the nominal price of oil and shows that there is no compelling evidence that survey forecasts outperform the no-change forecast. Gay, Simkins, & Turac (2009) analyzes how the market incorporates Bloomberg analysts' forecasts in the price of natural gas. Stark (2010) focuses on how accurate are the Survey of Professional Forecasters (SPF) predictions of inflation, unemployment, concluding that in general the SPF outperform the benchmark projections of univariate autoregressive time series at short horizons. Pierdzioch, Rülke, & Stadtmann (2010) find strong evidence of anti-herding among oil price forecasters. Cortazar, Kovacevic & Schwartz (2015) validates model predictions using expectation surveys. Cortazar, Millard, Ortega, & Schwartz (2019) use futures and expectation surveys to estimate a constant term structure of oil risk premiums while Cortazar, Liedtke, Ortega, & Schwartz (2018) uses the same information to estimate a time-varying term structure of oil premiums and what are the micro and macro variables that explain their variation.

In this paper we explore the usefulness of including a survey of analysts´ forecasts as input in an oil price forecasting model. We address the high volatility of expectation forecasts by including also futures price data, a Kalman filter to allow for observational error and a three factor stochastic model that allows for a time-varying risk premium term structure (Cortazar et al. 2018). The forecasting model is applied to ten years of crude oil prices from NYMEX and analyst forecasts from Bloomberg.

Results are very encouraging showing that the model performs much better than the no-change model, commonly used as a benchmark and better than many of the forecasting models from the literature. We conclude that analyst forecasts is a valuable source of input data that should be considered in future forecasting models.

This paper is organized as follows. Section 2 discusses how to measure performance in a forecasting model and what are the typical results found in the literature. Section 3 presents the futures and analysts' forecasts data that will be used in our model. Section 4 presents the proposed forecasting model. Section 5 presents the model results. Finally, section 6 concludes.

2. Measuring Forecasting Model Performance

2.1 Model Performance Metrics

The standard way of measuring the performance of a forecasting model is to compare it with a no-change or random walk model without drift that acts like its benchmark. This model assumes the forecast of a future spot oil price is the current spot price:

$$\hat{\mathrm{S}}_{t+h|t} = \mathrm{S}_{\mathrm{t}}$$

with S_t the current spot price and $\hat{\mathrm{S}}_{t+h|t}$ it's forecast for *h* periods ahead.

In what follows we present two performance metrics relative to the no-change benchmark.

2.1.1 Relative MSPE - Mean Squared Prediction Error

The Mean Squared Prediction Error (MSPE), for a specific horizon h, is defined as the average squared difference between the spot and the predicted prices:

$$MSPE = \frac{\sum_t \left(S_{t+h} - \hat{S}_{i,t+h|t}\right)^2}{n}$$

where S_{t+h} is the spot price for *h* periods ahead, $\hat{S}_{i,t+h|t}$ denotes the forecast of S_{t+h}, of model *i* using data through time t, and n represents the number of forecasts.

The Relative MSPE (Stock & Watson, 2004) is defined as:

$$Relative\ MSPE_i = \frac{\sum_t \left(S_{t+h} - \hat{S}_{i,t+h|t}\right)^2}{\sum_t \left(S_{t+h} - \hat{S}_{0,t+h|t}\right)^2}$$

where *i* is the forecasting model analyzed and *i*=0 corresponds to the no-change benchmark model.

Under this formulation, a Relative MSPE below one means that model *i's* forecasts are more accurate than the no-change forecasts.

2.1.2 Directional Accuracy

The Directional Accuracy refers to how well the forecasting model predicts if the spot prices are going up or down. Following Yao and Tan (2000), we define $a_{t,h}$ which takes a value of 1 if the model is successful in the direction of the forecast made in *t*, for *h* periods ahead, and zero otherwise.

$$a_{t,h} = \begin{cases} 1 & if\ (S_{t+h} - S_t)(\hat{S}_{t+h} - S_t) > 0 \\ 0 & otherwise \end{cases}$$

Then, the Success Ratio for a period T is computed as:

$$Success\ Ratio\ _T^h = \frac{\sum_{t \in T} a_{t,h}}{n}$$

A model that has a Success Ratio over 0.5 implies that their forecasts are better than those of the no-change model which is expected to correctly forecast the direction 50% of the time.

2.2 Performance Results of Forecasting Models from the Literature

In what follows we summarize the performance results of a set of alternative forecasting models that may represent a "typical" result found in the literature for this kind of models[1].

The main conclusion that can be drawn is that results are mixed and not always satisfactory. For example, for the 1 and 3 months horizon, even though the Relative MSPE ratio is 0.98 on average on the given set of models, the maximum value of this metric is 1.01 and 1.52 for each horizon. The Success Ratios are between 0.45 and 0.58, and the average for these two horizons is 0.55. For the 6 and 9 months horizon the Relative MSPE ratio ranges between 0.93 and 1.01 and the average is 1.02. On the other hand, Success Ratios are between 0.42 and 0.55. The average of this metric 0.49. If we look at the 12, 15 and 18 months horizon, the Relative MSPE ratio is between 0.92 and 1.11, while the average is 0.98. The average Success Ratio is 0.48 and it moves between 0.37 and 0.56.

[1] A more detailed discussion of these alternative models can be found in Appendix A

Finally, for the 21 and 24 months horizon, the Relative MSPE average is 1.01, while the Success Ratios goes from 0.35 and 0.49, with an average below 0.5.

From the above analysis it becomes clear that forecasting oil prices is a challenging issue, and explains why there are continuing efforts to improve the forecasting models.

3. Using Analyst Forecasts and Futures Prices Data

In this paper we propose using analyst expected price surveys, together with futures price data, to extract filtered price forecasts. We use Bloomberg's surveys for WTI's expected prices, a list of predictions done by different financial firms. The expectations, when given, are quarterly for the next 6 quarters and yearly, at most for the next 5 years. Weekly WTI oil futures prices quoted at NYMEX are also obtained from Bloomberg.

Most of the previous uses of survey data in the literature has been done using raw data, with no filtering or, at most, with a very simple averaging processing. This, we believe, may hinder the true value of this information. Data is available only when one of the many firms makes a prediction[2], and may be available any day of the week and for any horizon. Thus, the number of analyst's predictions is variable and the values volatile.

[2] Analyst forecasts are made for the average price on each quarter, or year, but following Cortazar et. el (2018) we assume they represent the price in the middle of their time period.

Figure 3.1 shows analysts´ forecasts and futures prices available up to a 24 month horizon, for the third week of September, 2015. It can be seen that expected price data is very volatile making a filtering procedure very valuable.

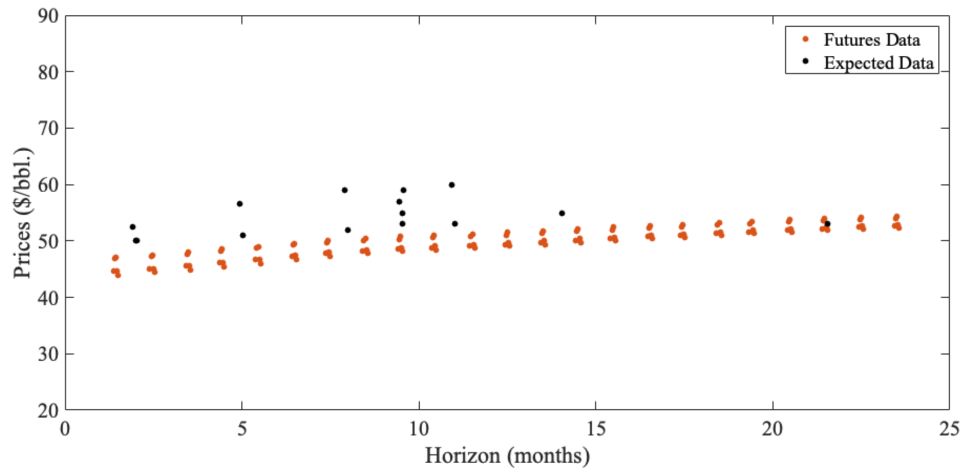

Figure 3.1: Futures and Analysts´ Expected Price Data, third week, September 2015.

Table 3.1 compares the amount of data for analysts expected prices and futures prices for a horizon of 24 months or less. It can be seen that there is much more futures data than analysts´ forecasts.

Table 3.1: Available Data of Futures and Bloomberg Expected Prices up to 24 months.

Year	Amount of data	
	Bloomberg's Analysts Expected Prices	Futures Prices
2007	375	5850
2008	639	5882
2009	512	5852
2010	374	5844
2011	445	5844
2012	632	5850
2013	824	5853
2014	1165	5846
2015	1207	5843
2016	1315	5849
2017	1205	5834
2018	1728	5852
Average	868.42	5849.92

Given the disparity in data availability between both sets of data we follow Cortazar et al. (2018) and use futures weekly data represented by each Wednesdays´ prices for maturities every 6 months (and the closest maturity forward) and analysts´ expected weekly data represented by the week’s average for each horizon.

Tables 3.2 and 3.3 summarize the weekly Analysts´ Expected and Futures Prices for horizons up to 24 months from 2007 to 2018.

Table 3.2: Expected prices between 2007 and 2018 for horizons up to 24 months.

Maturity (months)	Mean Price ($/bbl.)	Number of Observations
0 - 1	73.99	241
2 - 6	75.46	677
7 - 12	77.67	779
13 - 18	78.19	600
19 - 24	81.20	270
All maturities	77.23	2567

Table 3.3: Futures prices between 2007 and 2018 for horizons every 6 months up to 24 months.

Maturity (months)	Mean Price ($/bbl.)	Number of Observations
0 - 1	75.18	625
6	76.44	625
12	76.73	637
18	76.56	645
24	76.29	625
All maturities	76.24	3157

The forecasting model predictions will later be compared with the WTI spot price data available from the U.S, Energy Information Administration (EIA). Figure 3.2 presents the spot prices between 2007 and 2018.

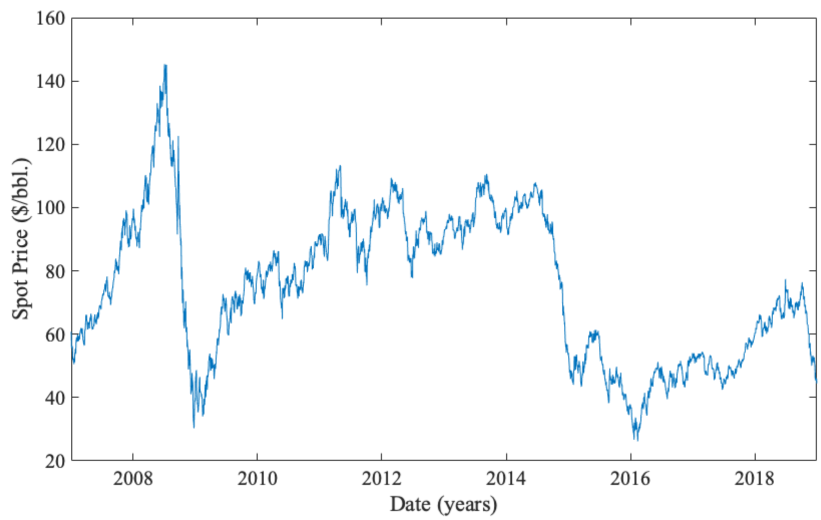

Figure 3.2: WTI Spot Prices between 2007 and 2018 (EIA - U.S. Energy Information Administration).

Figure 3.2 shows several sudden drops in oil prices being the major ones in the second half of 2008 and in the second half of 2014. This price behavior is going to be relevant when we analyze the forecasting performance of the model. Table 3.4 shows the annual average and standard deviations of WTI spot prices. It can be noticed the great variations between different years.

Table 3.4: Annual Average and Standard Deviation of WTI Spot Prices.

Year	Spot Price Average ($US/bbl.)	Spot Price Standard Deviation
2007	72.34	12.88
2008	99.67	28.62
2009	61.95	13.39
2010	79.48	5.25
2011	94.88	8.08
2012	94.05	7.73
2013	97.98	5.46
2014	93.17	13.55
2015	48.66	6.83
2016	43.29	6.74
2017	50.80	3.92
2018	65.23	6.53
All years	75.16	23.25

4. The Forecasting Model

4.1 Model Definition

As stated before, the proposed forecasting model should take into account the noisiness of the relatively few data in forecasting surveys, jointly with the more complete futures data.

We propose using the Cortazar et al. (2018) model, but instead of studying the risk premium behavior, as the authors do, analyze the model´s forecasting power.

In what follows we present the Cortazar et al. (2018) model, a non-stationary 3-factor stochastic model calibrated using both futures and analysts´ expectations. The model addresses the noisiness of the data by using a Kalman Filter, and generates the futures and expected price curves to extract the term structure of risk premiums.

Let S_t be the spot price at time t, then:

$$ln(S_t) = Y_t = h'x_t$$

$$dx_t = \left(-Ax_t + \begin{bmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right) dt + dw_t$$

where h is an $n\ x\ 1$ vector of constants, x_t is an $n\ x\ 1$ vector of state variables, b_1 is a scalar, A is an $n\ x\ n$ upper triangular matrix with its first diagonal element being zero the other diagonal elements all different and strictly positive. Let dw_t be an $n\ x\ 1$ vector of uncorrelated Brownian motions:

$$dw_t dw_t' = Idt$$

where I is an $n\ x\ n$ identity matrix.

Let RP_t be the risk premium at time t and assume the following expression:

$$RP_t = \lambda + \Lambda x_t$$

So, the risk adjusted version of the model is:

$$Y_t = h' x_t$$

$$dx_t = \left(-(A+\Lambda)x_t + \begin{bmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \lambda \right) dt + dw_t^Q$$

where λ is an $n\ x\ 1$ vector, Λ is an $n\ x\ n$ matrix that is not diagonal or triangular and dw_t^Q is a Brownian motion under risk neutral measure. Further restrictions on λ and Λ are not necessary.

Cortazar et al. (2018) shows that the futures price (define as the spot price S_t under the risk neutral measure Q), is given by the expression:

$$F_t(T) = E_t^Q(S_T) = e^{E_t^Q(Y_T)+\frac{1}{2}Var^Q(Y_T)}$$

with

$$E_t^Q(x_T) = e^{-(A+\Lambda)(T-t)}x_t + \left(\int_0^{T-t} e^{-(A+\Lambda)\tau} d\tau\right)(b-\lambda)$$

$$Cov^Q(x_T) = \int_0^{T-t} e^{-(A+\Lambda)\tau}(e^{-(A+\Lambda)\tau})' d\tau$$

The expected price satisfies the following equations:

$$E_t(S_T) = e^{E_t(Y_T)+\frac{1}{2}Var(Y_T)}$$

$$E_t(x_T) = e^{-A(T-t)}x_t + \left(\int_0^{T-t} e^{-A\tau} d\tau\right) b$$

$$Cov(x_T) = \int_0^{T-t} e^{-A\tau}(e^{-A\tau})' d\tau$$

Finally, model implicit volatilities of future prices σ_F and expected prices σ_E may be determined as follows:

$$\sigma_F = \sqrt{h' e^{-(A+\Lambda)(T-t)} e^{-(A+\Lambda)(T-t)'} h}$$

$$\sigma_E = \sqrt{h' e^{-A(T-t)} e^{-A(T-t)'} h}$$

4.2 Model Estimation

Cortazar et al. (2018) uses Kalman Filter (Kalman, 1960) to estimate parameters and state variables. At any given time-iteration, the Kalman Filter representation can be expressed by two equations. The first one is the following equation:

$$z_t = H_t x_t + d_t + v_t \quad v_t \sim N(0, R_t)$$

where z_t is an $m_t\ x\ 1$ vector that contains futures and expected log-prices observations at time t. H_t is a $m_t\ x\ n$ matrix, d_t is an $m_t\ x\ 1$ vector and v_t is a measurement error vector of $m_t\ x\ 1$ dimension with zero mean and covariance R_t. x_t is an $n\ x\ 1$ vector of state variables. In the model, m_t depends on the number of observations at each time, so the dimension of $z_t, H_t, d_t, v_t\ y\ R_t$ changes at each time.

Matrix R_t is defined by:

$$R_t = \begin{bmatrix} {\sigma_f}^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & {\sigma_f}^2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & {\sigma_e}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & {\sigma_e}^2 \end{bmatrix}$$

The second equation in the Kalman Filter is:

$$x_{t+1} = \bar{A}x_t + \bar{c} + w_t \quad w_t \sim N(0, Q)$$

where $\bar{A}$ is an $n\,x\,n$ matrix, and $\bar{c}$ is an $n\,x\,1$ vector. $\bar{A}$ and $\bar{c}$ represents the discretization of the process. In the above expression w_t is a vector of random variables with zero mean and an $n\ x\ n$ covariance matrix Q.

5. Model Results

5.1 Model Fit

Before presenting our forecasting results in this section, we now discuss the calibration process and the model errors. In order to be able to make out-of-sample forecasts for each year, we estimate the model parameters using data from all previous years, with a minimum of two years. Given that our data is from 2007 to 2018, our first forecast is for 2009 using data from 2007 to 2008, and our last one is for 2018 with data from 2007 to 2017.

Figure 5.1 shows the forecast and the futures curves for the third week of September of 2015 and the data available for that date[3]. The model considers the survey of expected prices and the futures data to jointly estimate both curves. Given the volatility of survey data, the Kalman filter optimally takes this into account. Thus the forecast curve does not

[3] The date is the same as in Figure 3.1, but now survey data for each horizon in the week has been averaged, and futures prices is only those of Wednesday (every 6 months).

necessarily perfectly fit current data but considers also past ones. Given that futures prices are less volatile, the futures curve fits much better the futures data.

Table 5.1 computes Mean Absolute Percentage Error (MAPE) for Analysts´ forecasts and Futures prices for each calibration year. As expected, the MAPE for the Futures data is much lower and less volatile than for the Analysts´ forecasts.

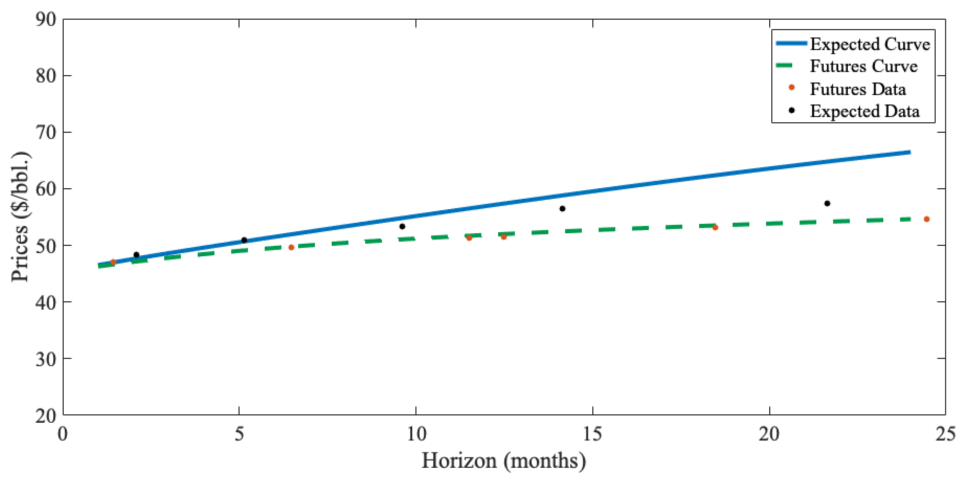

Figure 5.1: Forecast and Futures curves and Analysts´ Expected Price and Futures Data. Third week, September 2015.

Table 5.1: Mean Absolute Percentage Error (MAPE) of futures and forecast curves.

Calibration Years	Year	MAPE (%) between	
		Curve and Futures Prices Data	Curve and Analysts Expected Prices Data
2007-2008	2009	0.29%	9.63%
2007-2009	2010	0.38%	10.54%
2007-2010	2011	0.35%	10.35%
2007-2011	2012	0.37%	10.83%
2007-2012	2013	0.38%	11.09%
2007-2013	2014	0.36%	10.67%
2007-2014	2015	0.38%	9.97%
2007-2015	2016	0.41%	9.61%
2007-2016	2017	0.41%	9.45%
2007-2017	2018	0.42%	9.26%
Average		0.37%	10.14%
Standard Deviation		0.04%	0.64%

5.2 Forecasting Results of the Proposed Model

5.2.1 Relative MSPE

Table 5.2 summarizes the Relative MSPE results. They are shown for different forecasting horizons, for each year and for the whole 10 year period.

The two main conclusions that can be drawn from Table 5.2 are: First the general accuracy of the forecasting model over the whole 10 year period is much better than the no-change forecast. Second however good the model is when there is a relatively “reasonable” price behavior, this is not the case when there are sudden and unexpected jumps in spot prices, like those that can be seen in Figure 3.2. In these cases, analysts’ forecasts are obviously wrong and even the no-change model is a better alternative.

Table 5.2: Relative MSPE Forecasting Results.

Horizon (months)	General Accuracy	Yearly Accuracy									
		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	**0.995**	1.011	1.247	**0.980**	1.048	1.166	**0.785**	**0.981**	**0.843**	**0.906**	1.093
3	**0.954**	**0.753**	1.273	1.074	**0.998**	**0.855**	**0.882**	1.109	**0.524**	**0.937**	1.207
6	**0.871**	**0.531**	1.049	**0.857**	**0.719**	**0.944**	**0.805**	1.690	**0.432**	**0.899**	1.159
9	**0.826**	**0.343**	**0.935**	1.113	**0.893**	3.003	**0.726**	2.143	**0.471**	**0.778**	1.190
12	**0.788**	**0.238**	**0.904**	**0.764**	1.221	**0.746**	**0.749**	3.474	**0.382**	**0.663**	**0.957**
15	**0.710**	**0.254**	**0.823**	**0.571**	1.335	**0.539**	**0.767**	4.020	**0.231**	**0.564**	**0.814**
18	**0.671**	**0.371**	**0.589**	**0.543**	2.336	**0.452**	**0.758**	7.157	**0.107**	**0.357**	**0.403**
21	**0.676**	**0.338**	**0.653**	1.486	1.940	**0.557**	**0.763**	10.100	**0.083**	**0.366**	-
24	**0.651**	**0.297**	**0.532**	1.069	1.515	**0.590**	**0.775**	4.483	**0.143**	**0.465**	-
Horizons up to 12 months	**0.848**	**0.417**	**0.987**	**0.962**	**0.951**	1.038	**0.761**	2.040	**0.453**	**0.783**	1.132
Horizons up to 24 months	**0.729**	**0.346**	**0.800**	**0.929**	1.369	**0.565**	**0.764**	3.269	**0.189**	**0.643**	1.068

Boldface indicates improvements on the no-change forecast. The yearly accuracy refers to the year in which the forecast is made. The general accuracy indicates Relative MSPE from 2009 to 2018. "Horizons up to 12 months" indicates the Relative MSPE for horizons from 1 to 12 months. "Horizons up to 24 months" indicates the Relative MSPE for horizons from 1 to 24 months.

5.2.2 Directional Accuracy

Table 5.3 summarizes the Success Ratios. They are shown for different forecasting horizons, for each year and for the whole 10 year period.

Conclusions drawn from analyzing this table are similar, but even a bit stronger, than those from Table 5.2. Again our forecasting model is better than the no-change model over all horizons, when the 10 year period is considered. Moreover, this difference is statistically significant for all horizons at very high significance levels. Also our second conclusion holds again: when sudden jumps in spot prices occur, analysts´ forecasts may be even worse than those of the no-change model.

Table 5.3: Directional Accuracy Forecasting Results.

Horizon (months)	General Accuracy	Yearly Accuracy									
		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	**0.555** ***	0.500	0.385	**0.558** **	0.500	0.442	**0.830** ***	**0.538**	**0.635**	**0.635**	**0.519**
3	**0.541** **	**0.673** ***	0.308	**0.519** **	0.500	**0.519**	**0.679**	0.404	**0.692**	**0.635**	0.481
6	**0.580** ***	**0.769** ***	0.404	**0.519** *	**0.577** **	**0.538** ***	**0.868** ***	0.269	**0.808**	**0.750**	0.288
9	**0.635** ***	**0.769**	**0.519** **	**0.635** ***	**0.558** **	0.462	**0.849**	0.327	**0.769**	**0.942**	**0.510**
12	**0.692** ***	**1.000** ***	**0.596**	**0.731** ***	0.404*	**0.577** *	**0.830**	0.346	**0.846**	**0.904** ***	**0.684**
15	**0.733** ***	**1.000** ***	**0.673**	**0.808** ***	0.365	**0.712** ***	**0.811**	**0.519**	**0.923**	**0.808**	**0.680**
18	**0.775** ***	**1.000** ***	**0.865** ***	**0.750** ***	0.212	**0.923** ***	**0.792**	0.481	**1.000** ***	**0.904**	**1.000** ***
21	**0.752** ***	**0.981**	**0.712**	**0.615** **	0.346	**0.962**	**0.774**	**0.615**	**1.000** ***	**0.765**	-
24	**0.818** ***	**1.000** ***	**0.827** ***	**0.615** ***	**0.635** **	**0.962**	**0.774** **	**0.673**	**0.962**	**0.947**	-
Horizons up to 12 months	**0.602** ***	**0.760** ***	0.447***	**0.575** ***	**0.527** ***	**0.537** ***	**0.807** ***	0.335	**0.771**	**0.787**	0.461
Horizons up to 24 months	**0.676** ***	**0.877** ***	**0.591** ***	**0.638** ***	0.455***	**0.694** ***	**0.800** ***	0.435	**0.867**	**0.816**	**0.526** **

Boldface indicates improvements on the no-change forecast. The yearly accuracy refers to the year in which the forecast is made. Statistical significance levels are given by ***1%, **5% and *10%, according to the Pesaran-Timmermann test (2009). The general accuracy indicates Directional Accuracy from 2009 to 2018. "Horizons up to 12 months" indicates the Directional Accuracy for horizons from 1 to 12 months. "Horizons up to 24 months" indicates the Directional Accuracy for horizons from 1 to 24 months.

5.2.3 Comparing Forecasting Results with those of Alternative Models

Even though our accuracy results are quite promising, in this section we want to make a quick comparison with some results published from other models in the literature. As stated in Section 2.2 we chose 6 alternative models from the literature[4] to analyze their results, but some caveats should be taken into account.

[4] See Appendix A for a brief description of each of the 6 alternative models

First, the selection of the 6 models with which our model will be compared is somewhat arbitrary and does not necessarily represents the best models from the literature.

Second, given the data requirements for the estimation of the expected curve, the amount of out-of-sample data that of our model is only 10 years which is sometimes less than those used by other models.

Third, the model developed in this paper predicts nominal prices, while there is another type of models that is dedicated to predicting real prices. Although given that we are comparing forecasts for a maximum of two years this may not be a great concern.

Finally the evaluation period of each model is different across the papers, so depending on the behavior of the spot prices during the forecasting period comparisons could be unfair.

Despite these elements, the objective of this section is not to determine the best model, but rather to see if the results are reasonably close to those of other models thus to determine if there is valuable information in the analysts´ expected prices.

Table 5.4 compares the results of our model with those of the alternative models. It can be seen that our model has an excellent performance compared with the alternative 6 models. In terms of the Relative MSPE metrics our forecast is better than all the other models for all horizons, except for the one and three month periods. In terms of the Directional Accuracy metric, results of our model are even better. Our model overcomes all others for all horizons except for the 3 month forecasts.

As stated before, by showing our good results we are not claiming our model is the best in the literature, but only that using analysts´ forecasts filtered in a proper way, seems to be a data source that should not be discarded.

Table 5.4: Results of different Forecasting models.

Horizon (months)	Error metric	Our model	Alternative models		
			Average	Minimum	Maximum
1	Relative MSPE	**0.995**	**0.985**	**0.917**	1.008
	Directional Accuracy	**0.555**	**0.505**	0.460	**0.554**
3	Relative MSPE	**0.954**	1.077	**0.936**	1.519
	Directional Accuracy	**0.541**	0.486	0.447	**0.575**
6	Relative MSPE	**0.871**	1.032	**0.935**	1.040
	Directional Accuracy	**0.580**	**0.504**	0.459	**0.541**
9	Relative MSPE	**0.826**	1.002	**0.927**	1.099
	Directional Accuracy	**0.635**	0.481	0.419	**0.553**
12	Relative MSPE	**0.788**	**0.986**	**0.917**	1.112
	Directional Accuracy	**0.692**	**0.515**	0.370	**0.557**
15	Relative MSPE	**0.710**	**0.984**	**0.936**	1.031
	Directional Accuracy	**0.733**	0.464	0.434	0.494
18	Relative MSPE	**0.671**	**0.984**	**0.969**	1.041
	Directional Accuracy	**0.775**	0.474	0.397	0.440
21	Relative MSPE	**0.676**	1.023	**0.987**	1.058
	Directional Accuracy	**0.752**	0.393	0.349	0.437
24	Relative MSPE	**0.651**	**0.997**	**0.940**	1.054
	Directional Accuracy	**0.818**	0.429	0.367	0.491

Boldface indicates improvements on the no-change forecast. “Our Model” column contains the general accuracy results for the indicated horizon. The “Average” column indicates the simple average between the other models’ results when reported their error metrics for the indicated horizon. The “Minimum” and “Maximum” columns indicate the minimum and maximum value of the indicated error metric from the other models. Other models refers to the models reviewed in Appendix A.

6. Conclusions

We have argued that even though there is a wide consensus that having good oil price forecasts is very valuable for many agents in the economy, results have not been fully satisfactory. There have been many attempts to improve this result, from new models to new inputs. Research has explored many different modeling approaches including time series, regressions, and artificial intelligence. On the other hand, many different sources of input data have been used like spot and futures prices, product spreads, and micro and macro variables.

In this paper we explore the use of analysts´ expected prices from different financial firms and surveyed by Bloomberg. The proposed model uses this information, together with futures prices, in a multifactor stochastic model calibrated with the Kalman filter.

We discuss that performance of forecasting models are typically measured using the Relative MSPE and Directional Accuracy and compared with a no-change forecast benchmark. Using these metrics our results are very encouraging when compared to other models from the literature.

Our main conclusion is that using analysts´ expected prices, even if they are noisy, is a useful data source that should be taken into account in the future by commodity price forecasters.

REFERENCES

Alquist, R., Kilian, L., & Vigfusson, R. (2013). Forecasting the price of oil. In In Handbook of economic forecasting. (Vol. 2, pp. 427-507). Elsevier.

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

Baumeister, C., & Kilian, L. (2012). Real-Time Forecasts of the Real Price of Oil. Journal of Business & Economic Statistics, 30(2), 326-336.

Baumeister, C., Kilian, L., & Lee, T. (2014). Are there gains from pooling real-time oil price forecasts? Energy Economics, 46, S33–S43.

Baumeister, C., Kilian, L., & Zhou, X. (2018). Are product spreads useful for forecasting oil prices? An empirical evaluation of the Verleger hypothesis. Macroeconomic Dynamics, 22(3), 562-580.

Busse, M., Knittel, C., & Zettelmeyer, F. (2009). Pain at the pump: how gasoline prices affect automobile purchasing in new and used markets. NBER working paper, 15590.

Cheng, F., Li, T., Wei, Y., & Fan, T. (2019). The VEC-NAR model for short-term forecasting of oil prices. Energy Economics, 78, 656-667.

Cortazar, Gonzalo, Ivo Kovacevic, and Eduardo S. Schwartz, 2015, Expected commodity returns and pricing models, Energy Economics 49, 60–71

Cortazar, G., Liedtke, P., Ortega, H., & Schwartz, E. (2018). Time-Varying Term Structure of Oil Risk Premiums. Working Paper.

Cortazar, G., Millard, C., Ortega, H., & Schwartz, E. (2019). Commodity Price Forecasts, Futures Prices and Pricing Models. Management Science. Vol. 65 (9) 4141-4155

Gay, G., Simkins, B., & Turac, M. (2009). Analyst forecasts and price discovery in futures markets: The case of natural gas storage. The Journal of Futures Markets, 29(5), 451–477.

Kilian, L., & Murphy, D. (2014). The role of inventories and speculative trading in the global market for crude oil. Journal of Applied Economics, 29, 454-478.

Lanza, A., Manera, M., & Giovannini, M. (2005). Modeling and forecasting cointegrated relationships among heavy oil and product prices. Energy Economics, 27, 831-848.

Li, J., Zhu, S., & Wu, Q. (2019). Monthly crude oil spot price forecasting using variational mode decomposition. Energy Economics, 83, 240-253.

Miao, H., Ramchander, S., Wang, T., & Yang, D. (2017). Influential factors in crude oil price forecasting. Energy Economics, 68, 77-88.

Pesaran, M., & Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. Journal of the American Statistical Association (104), 325-337.

Pierdzioch, C., Rülke, J. C., & Stadtmann, G. (2010). New evidence of anti-herding of oil-price forecasters. Energy Economics, 32(6), 1456-1459.

Stark, T. (2010). Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters. Federal Reserve Bank of Philadelphia Research Rap, Special Report, 1.

Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. Journal of forecasting, 23(6), 405-430.

Sun, S., Sun, Y., Wang, S., & Wei, Y. (2018). Interval decomposition ensemble approach for crude oil price forecasting. Energy Economics, 76, 274-287.

Wang, Y., Liu, L., Diao, X., & Wu, C. (2015). Forecasting the real prices of crude oil under economic and statistical constraints. Energy Economics, 51, 599-608.

Yao, J., & Tan, C. L. (2000). A case study on using neural networks to perform technical forecasting of forex. Neurocomputing, 34(1-4), 79-98.

Yu, L., Wang, S., & Lai, K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics, 30(5), 2623-2635.

Zhang, J. L., Zhang, Y. J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. Energy Economics(49), 649-659.

Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. Energy Economics, 66, 9-16.

APPENDIX

Appendix A: Alternative forecasting models

In the following Appendix, we present 6 different forecasting models from the literature.

A.1 Single product spot spread model

Inspired by the idea that gasoline prices and crude oil prices move together in the long run (Lanza et al., 2005), Baumeister et al. (2018) developed a single product spread model to forecast the real WTI price of oil. In what follows, there is a brief review of that model and its results.

A.1.1 Data

The forecast evaluation period is from early 1992 until September 2012, that means more than twenty years of out-of-sample data evaluation. They use data from WTI, gasoline and heating oil spot prices and, a proxy for the inflation rate.

A.1.2 Model

The single product spread model is:

$$\widehat{S}_{t+h|t}{}^{oil} = S_t^{oil}\, exp\left[\widehat{\alpha} + \widehat{\beta}\left(p_t^i - p_t{}^{oil}\right) - \boldsymbol{E}\left(\pi_{t+h}^h\right)\right]$$

Where $i \in \{gasoline, heating\ oil\}$, $\widehat{S}_{t+h|t}{}^{oil}$ denote the forecast of oil for h periods ahead, using data through time t. S_t^{oil} is the current real (t) price of oil, α and β

are estimated parameters and p_t^i is the current log-price of i. ${p_t}^{oil}$ is the current log-price of oil and $\boldsymbol{E}\left(\pi_{t+h}^h\right)$ denotes the proxy for the inflation rate from t to $t+h$.

A.1.3 Results

The forecast accuracy of this model is evaluated using both Relative MSPEs relative to no-change forecast and Success Ratios. On their analysis, better results are obtained fixing $\alpha = 0$. Results are shown in Table A.1.

Table A. 1: Forecast accuracy of single product spot spread models for the real WTI price developed by Baumeister et al. (2018).

Horizon (months)	Error metric	Model	
		Gasoline spot spread	Heating oil spot spread
h=1	Relative MSPE	**0.999**	1.008
	Success Ratio	**0.554**	**0.534**
h=3	Relative MSPE	**0.998**	1.023
	Success Ratio	**0.575**	0.482
h=6	Relative MSPE	**0.978***	1.037
	Success Ratio	**0.541**	0.459
h=9	Relative MSPE	**0.965****	1.052
	Success Ratio	0.419	0.419
h=12	Relative MSPE	**0.940****	1.040
	Success Ratio	**0.504**	0.370
h=15	Relative MSPE	**0.936****	1.031
	Success Ratio	0.494	0.434
h=18	Relative MSPE	**0.969***	1.041
	Success Ratio	0.440	0.397
h=21	Relative MSPE	**0.987**	1.058
	Success Ratio	0.437	0.349
h=24	Relative MSPE	**0.940****	1.054
	Success Ratio	0.491	0.367

Boldface indicates improvements on the no-change forecast. Statistically significant reductions in the MSPE according to the Clark-West (2007) test and statistically significant improvements in Directional Accuracy according to the Pesaran-Timmermann test (2009) are marked using * (10% significance level) and ** (5% significance level) respectively.

This paper concludes that the most accurate single spread forecasting model is a model based on the gasoline spot spread alone. From Table A.1 it can be observed that for the gasoline spot spread model, all horizons have a Relative MSPE lower than 1. This does not happen with the heating oil spot spread model: all the horizons present Relative MSPEs above 1. Regarding the Success Ratios, for the gasoline spot spread model there are 4 out of 9 horizons with Success Ratios above 0.5, this means that 44% of the monthly horizons are more accurate than flipping a coin. On the other hand, the heating oil spot spread model only has one horizon with a Success Ratio over 0.5.

A.2 EMD-based neural network model

As anticipated, inspired by the “divide-and-conquer” principle, Yu et al. (2008) develop an empirical mode decomposition (EMD) based on neural networks ensemble learning method. They also compare its prediction with other forecasting model (ARIMA) by using error metrics such as the Directional Accuracy. They do this work for both WTI and Brent crude oil spot prices.

A.2.1 Data

In order to train their model, they use information from January 1986 to December 2000, and, to evaluate the model performance, data from January 2001 to September 2006. This is 6 years in out-of-sample analysis.

A.2.2 Model

This model is actually an “EMD-FNN-ALNN” ensemble learning approach. That is, it is an “EMD (Decomposition)–FNN (Prediction)– ALNN (Ensemble)” methodology. The EMD technique decompose a time series $x(t)$ into a sum of oscillatory functions, namely intrinsic mode functions (IMFs). In order to obtains the IMFs, they decompose the data series using a standard procedure until the spot criterium is satisfied and, at the end of this procedure, the data series can be expressed by:

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t)$$

Where n is the number of IMFs, $r_n(t)$ is the final residue, representing the central tendency of the data series. The residue is calculated on every step as the difference between $x(t)$ and the current IMF. $c_j(t)$ are the IMFs – which are nearly orthogonal to each other, and all have nearly zero means (Yu et al., 2008).

A.2.3 Results

Although in this work a horizon-analysis is not performed, they present their overall results. Regarding Directional Accuracy, they show a Success Ratio of 0.87 for WTI prices, a very good result compared with the results of their benchmark (ARIMA model got 0.52).

A.3 ARMA, oil futures and commodity price-based models

Baumeister and Kilian (2012) explore the forecast accuracy of five different models, including an autoregressive moving average (ARMA). Also, they present alternative forecasting methods, such as the oil futures-based and the commodity price-based model.

In what follows, the data used, the model specifications and their results are described.

A.3.1 Data

The evaluation window in this paper is from January 1992 to June 2010, this means that there are 18.5 years to analyze. The ARMA model uses information from the past

WTI spot prices in order to forecast the real price. While the futures-based model uses information from the futures prices, the commodity price-based model uses information from raw industrial materials.

A.3.2 Model

The ARMA(1,1) model for X_t can be expressed as follows:

$$X_t = \mu + \phi X_{t-1} + \epsilon_t + \theta\epsilon_{t-1} \qquad \forall t$$

$$\Phi(L)X_t = \mu + \Theta(L)\epsilon_t$$

Where $\theta \neq 0$, $\phi \neq 0$, μ is a constant term and L is the lag operator, ϵ_t is a weak while noise process with mean zero and variance given by σ_ϵ^2. $\Phi(L) = 1 - \phi L$ and $\Phi(L) = 1 - \theta L$.

Adapting this to a forecasting model involves that X_t represents the forecast for the price of oil given by this model. This model is estimated numerically by Gaussian maximum likelihood methods.

The oil futures-based model:

$$\widehat{S}_{t+h|t}{}^{oil} = S_t^{oil}\left(1 + f_t^h - s_t - \boldsymbol{E}\left(\pi_{t+h}^h\right)\right)$$

$\widehat{S}_{t+h|t}{}^{oil}$ denote the forecast of oil for h periods ahead, using data through time t. R_t^{oil} is the current real price of oil. f_t^h is the log of the current WTI oil futures price for maturity h, s_t is the log of the corresponding WTI spot price and $\boldsymbol{E}\left(\pi_{t+h}^h\right)$ is the expected U.S. inflation rate over the next h periods.

On the other hand, the commodity price-based model can be expressed:

$$\widehat{S}_{t+h|t}{}^{oil} = S_t^{oil}\left(1 + \pi_t{}^{h,industrial\ raw\ materials} - \boldsymbol{E}\left(\pi_{t+h}^h\right)\right)$$

Where $\pi_t{}^{h,industrial\ raw\ materials}$ stands for the percent change in the Commodity Research Bureau (CRB) index of the spot price of industrial raw materials (other than oil) over the preceding h months. The term $\boldsymbol{E}(\pi_{t+h}^{h})$ is the expected U.S. inflation rate over the next h periods.

A.3.3 Results

The results presented in this paper for the ARMA(1,1), oil futures-based and commodity price-based models are detailed in Table A.2.

Table A. 2: Forecast accuracy for oil futures-based and commodity price-based models for the real WTI price (Baumeister & Kilian, 2012).

Horizon (months)	Error metric	Model		
		ARMA(1.1)	Oil futures-based model	Commodity price-based model
h=1	Relative MSPE	**0.917****	1.004	**0.820**
	Success Ratio	0.496	0.460	**0.550***
h=3	Relative MSPE	**0.936****	**0.999**	**0.744***
	Success Ratio	0.482	0.477	**0.609****
h=6	Relative MSPE	**0.935****	1.002	1.040
	Success Ratio	0.465	**0.502**	**0.590****
h=9	Relative MSPE	**0.927****	**0.981**	1.099
	Success Ratio	0.467	**0.547****	**0.565***
h=12	Relative MSPE	**0.917****	**0.932**	1.112
	Success Ratio	**0.507**	**0.559****	**0.574***

Boldface indicates improvements on the no-change forecast. For the ARMA model, ** denotes significance at the 5% level based on bootstrap critical values for the average loss differential. For the oil futures-based and the commodity price-based models, ** denotes significance at the 5% level and * at the 10% level, based on the Diebold and Mariano (1995) test of equal predictive accuracy and the Pesaran-Timmermann test (2009) test of the null hypotesis of no Directional Accuracy.

Based on the 5 months horizon in analysis, the Relative MSPE shows that this model has better results than the no-change forecast, but, Success Ratios do not show a good performance. Only one out of five monthly horizon Success Ratios is above 0.5 (12 months horizon).

The oil futures-based model shows relatively good results for the 9 and 12 months horizons (both in Relative MSPE and Directional Accuracy metrics), compared with the 1, 3 and 6 months horizon. On the other hand, commodity price-based model shows good performance for the 1 and 3 months horizons (from both error metrics), while for the other horizons in analysis there are good results only for Directional Accuracy (Relative MSPEs are above 1).

A.4 Forecast methods based on monthly future prices and surveys

Alquist et al. (2013) performs a vast review on how to forecast the crude oil price. Within this work, models of both real prices and nominal, and for both short and long horizons are reviewed. Among the forecasts associated with nominal and short-horizon prices, models based on monthly future prices and those that take the surveys forecasts of this nominal price of WTI are found.

A.4.1 Data

For the methods based on monthly future prices the forecast evaluation period is from January 1991 to December 2009, this means 19 years of data.

On the other hand, the survey based forecasts uses another set of data, that is different for every forecast. The fist source is Consensus Economics (CE), and the evaluation period on this forecast is from October 1989 to December 2009, this means more than 20 years of data.

A.4.2 Model

This work presents 18 different forecast methods based on monthly future prices. Following it shows two of them: simple monthly-futures and the monthly-futures spread model

Simple monthly-futures model	$\hat{S}_{t+h\|t}{}^{oil} = F_t^h$
Monthly-futures spread model	$\hat{S}_{t+h\|t}{}^{oil} = S_t^{oil}\left(1 + \hat{\beta}\, ln\left(\frac{F_t^h}{S_t^{oil}}\right)\right)$

$\hat{S}_{t+h|t}{}^{oil}$ denote the forecast of oil for h periods ahead, using data through time t. S_t^{oil} is the current real price of oil. F_t^h is the current WTI oil futures price for maturity h and $\hat{\beta}$ it's an estimated parameter from the model.

Regarding the forecasts methods based on the CE survey, they use the arithmetic mean at their relevant horizons (three and twelve months):

$$\hat{S}_{t+h|t}{}^{oil} = S_{t,h}^{CE}$$

Where $S_{t,h}^{CE}$ denotes the arithmetic mean of the forecasts made by the firms for the specified period t, for h periods ahead.

A.4.3 Results

Using the described models, they report the Relative MSPE and the Success Ratios for horizons up to 12 months.

Table A. 3: Forecast accuracy for methods based on monthly future prices and surveys for the nominal WTI price (Alquist et al., 2013).

Horizon (months)	Error metric	Model		
		Simple monthly-futures	Monthly-futures spread	CE Survey based
h=1	Relative MSPE	**0.988**	**0.995**	
	Success Ratio	0.465	**0.531***	
h=3	Relative MSPE	**0.998**	**0.990**	1.519
	Success Ratio	0.465	0.474	0.447
h=6	Relative MSPE	**0.991**	**0.978**	
	Success Ratio	**0.509**	**0.535**	
h=9	Relative MSPE	**0.978**	**0.989**	
	Success Ratio	**0.548**	**0.553**	
h=12	Relative MSPE	**0.941**	1.052	**0.944**
	Success Ratio	**0.557***	**0.528**	**0.539***

Boldface indicates improvements on the no-change forecast. ** denotes significance at the 5% level and * at the 10% level, based on the Diebold and Mariano (1995) test of equal predictive accuracy and the Pesaran-Timmermann test (2009) test of the null hypotesis of no Directional Accuracy.

Regarding the Relative MSPE, between the first two models under analysis (associated with monthly futures prices), we can see that the simple model of monthly futures has slightly better results for the horizons in analysis, being - for both models - this ratio less than 1 (except for the second model, within 12 months). Meanwhile, regarding the success rate, there is no clear superiority of either, but from 6 months on it is above 0.5. This, unlike the horizons equal to 1 and 3 months, where the only favorable result is for the second model for the one month horizon.

Regarding the performance of the CE survey based model in analysis, the results for the 3 month-horizon is quite unfavorable in both Relative MSPE and Directional

Accuracy. This, unlike the results obtained for the 12 month-horizon, where both metrics reflect a good performance compared to the no-change forecast.

A.5 Constrained Predictive Regression

Using regressions of economic variables to predict the price of oil is very usual, but the interesting thing about this work (Wang et al., 2015) lies in the use of three restrictions according to the analysis of in-sample relations between real oil prices and its predictors. The first restriction is based on the economic theory while the second one is based on the statistical significance of regression coefficients. The third restriction is the combination of both economic and statistical restrictions.

A.5.1 Data

The sample data span the period from July 1986 through July 2013. Of this, the initial sample period covers the first 5-years, while the out-of-sample period covers the remaining 265 months. This represents 22 years of evaluation data.

A.5.2 Model

The standard predictive regression model can be written as:

$$y_{t+h} = \alpha_i + \beta_i * x_{i,t} + \theta_i * y_t + \varepsilon_{t+h}$$

Where y_{t+h} is the real oil price, h is the forecasting horizon, $x_{i,t}$ is an independent variable and ε_{t+h} is the disturbance term with the standard Gaussian distribution. The out-of-sample forecast is given by:

$$\hat{y}_{t+h} = \widehat{\alpha_\iota} + \hat{\beta}_i * x_{i,t} + \hat{\theta}_i * y_t$$

Where $\hat{y}_{t+h}$ is the real oil price forecast, $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\theta}_i$ are the ordinary least squared (OLS) estimates of α_i, β_i and θ_i, respectively.

They use 9 independent variables to predict: oil futures prices, global oil production, global economic activity, changes of oil inventory, US oil imports, US petroleum consumption, crack spread, non-energy commodity index and a speculative index.

A.5.3 Results

Table A.4 reports both Relative MSPE and Directional Accuracy out-of-sample results for 1, 6, 12 and 18 months horizons.

Table A. 4: : Forecast accuracy of the constrained regression model developed by Wang et al. (2015) for the real price of WTI.

Horizon (months)	Error metric	Constrained regression model
h=1	Relative MSPE	**0.985**
	Success Ratio	0.496
h=6	Relative MSPE	1.292
	Success Ratio	**0.515**
h=12	Relative MSPE	**0.943**
	Success Ratio	**0.598*** **
h=18	Relative MSPE	**0.941**
	Success Ratio	**0.585*** **

Boldface indicates improvements on the no-change forecast. *** denotes significance at the 1% level based on Pesaran-Timmermann test (2009) test of the null hypotesis of no Directional Accuracy.

The model outperforms the no-change forecast on two of the four in-analysis horizons in terms of both error metrics (12 and 18 months horizons). For the one-month horizon, the constrained regression model report a Relative MSPE ratio below 1, but the Success Ratio is below the desirable level. Regarding the six months horizon, its Relative MSPE is above 1, but the Success Ratio is above 0.5 for this horizon.

A.6. LASSO regression method

The LASSO model (Miao et al., 2017) is an innovative variable selection method and has found application in the energy area (electricity consumption, electricity prices and natural gas prices). The model uses 26 potential predictors that are classified into six groups: supply factors (global crude oil production, OPEC surplus crude oil production capacity, US crude oil closing stock, among others), demand factors (growth rate of China real GDP, world steel production, ISM manufacturing index, among others), financial factors (three month U.S. Treasury bill rate, S&P 500 Index, among others), commodity markets factors (S&P GSCI Non-Energy index and CRB Raw Materials Index), speculative factors (ratio of trading volume of oil futures contracts to global oil production) and political factors (total amount of terrorist attack in the Middle East and North Africa).

A.6.1 Data

The data spans the period from January 2002 to September 2015, and the model is estimated over various fixed length rolling windows (5, 6 and 7 years), we will present the 5-years rolling window results. They do not provide an out-of-sample analysis, so we are not going to use their results as a benchmark.

A.6.2 Model

The LASSO model selects the variables by adding a penalty term to the cost function, allowing the estimated value of the regression coefficients small. Once the variables are selected, the model minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant (c), which controls the degree of shrinkage that is applied to the estimates:

$$\widehat{\beta^L} = argmin\left\{\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{i=1}^{n}\beta_{i,j} * x_{i,j}\right)^2\right\}$$

Subject to:

$$\sum_{j=1}^{p}|\widehat{\beta_j^L}| \leq c \ (Constant)$$

In order to obtain a pre-determined sequence of LASSO solutions, they use a computationally efficient method, all the other LASSO solutions are obtained by linear interpolation form the sequence of LASSO solutions.

A.6.3 Results

Table A.5 reports both Relative MSPE and Directional Accuracy in-sample results for 1, 2, 4 and 8 months horizons.

Table A. 5: Forecast accuracy of the LASSO regression model developed by Miao et al. (2017). In sample results for the 5-years rolling window.

Horizon (months)	Error metric	LASSO
h=1	Relative MSPE	**0.924**
	Success Ratio	**0.572**
h=2	Relative MSPE	**0.911**
	Success Ratio	**0.533**
h=4	Relative MSPE	**0.896**
	Success Ratio	**0.544**
h=8	Relative MSPE	**0.873**
	Success Ratio	**0.562**

Boldface indicates improvements on the no-change forecast. Statistically significant reductions in the MSPE according to the Diebold and Mariano (1995) test and statistically significant improvements in Directional Accuracy according to the Pesaran-Timmermann test (2009) are marked using * (10% significance level), ** (5% significance level) and *** (1% significance level) respectively.

Relative to the no-change and futures-based models, LASSO forecasts at the 8-step ahead horizon yield significant reductions in MSPE with a ratio of 0.87, while the most impressive Success Ratio is above 0.57 (one-month horizon).

Appendix B: Parameters

In the following Appendix, we present the parameters obtained after the calibration process and used to make forecasts using the indicated years of data.

Table B. 1: Parameter estimates for the 3-factor model.

Parameter	Calibration Data Years				
	2007-2008	2007-2009	2007-2010	2007-2011	2007-2012
A_{11}	0.000	0.000	0.000	0.000	0.000
A_{12}	-0.043	-0.093	1.441	1.414**	1.391**
A_{13}	-0.010	0.565	0.566	0.595	0.606
A_{22}	1.919***	0.244*	3.831*	4.406**	3.987**
A_{23}	1.190	0.214	-0.265	0.612	0.648
A_{33}	0.241**	1.718***	1.082***	0.583***	0.559***
Λ_{11}	-0.018***	-0.002	0.000	-0.009***	-0.014***
Λ_{12}	-0.024	0.067	-1.527	-1.374**	-1.307**
Λ_{13}	-0.006	-0.642	-0.662	-0.582	-0.573
Λ_{21}	-0.056**	-0.034*	-0.011	0.002	0.000
Λ_{22}	-1.077***	-0.035	-3.089	-3.125	-2.710
Λ_{23}	-0.785	-0.420	0.744	-0.130	-0.173
Λ_{31}	-0.055*	-0.036***	-0.022	-0.003	0.006
Λ_{32}	0.643***	0.010	0.532***	0.347	0.329
Λ_{33}	0.444	-0.698***	-0.118	-0.279	-0.275
h_1	0.13***	0.122***	0.29***	0.258***	0.23***
h_2	0.694***	0.252***	0.008	0.323***	0.309***
h_3	0.178	0.711***	0.416***	0.409***	0.385***
λ_1	-0.061	-0.047	-0.017	0.015	0.049
λ_2	2.682***	1.5*	-0.111	-0.306	-0.317
λ_3	1.515	1.476***	0.172	0.162	0.045
b_1	-0.713**	-0.135	-0.18***	-0.195*	-0.247*
σ_f	0.006***	0.007***	0.005***	0.007***	0.006***
σ_e	0.129***	0.126***	0.123***	0.143***	0.143***

Table B.1 (cont.)

Parameter	Calibration Data Years				
	2007-2013	2007-2014	2007-2015	2007-2016	2007-2017
A_{11}	0.000	0.000	0.000	0.000	0.000
A_{12}	1.718***	0.098	-0.049	-0.116	1.986**
A_{13}	0.623	-0.043	-0.132	3.634**	0.409
A_{22}	3.061**	0.710	0.331*	0.43***	2.276***
A_{23}	0.570	1.499***	1.674***	-0.351	0.609
A_{33}	0.544**	0.349	1.45***	2.158***	0.748***
Λ_{11}	-0.015***	0.000	0.012	-0.001	-0.027
Λ_{12}	-1.616***	-0.097	-0.004	0.067	-2.008**
Λ_{13}	-0.581	0.184	0.237	-3.579**	-0.401
Λ_{21}	-0.014	-0.063***	-0.082***	-0.07***	-0.169***
Λ_{22}	-1.792	-0.303**	-0.146	-0.221*	-1.097
Λ_{23}	-0.105	-1.001	-1.647***	0.361	-0.112
Λ_{31}	0.014	-0.007	-0.063***	-0.095***	0.011
Λ_{32}	0.292	0.473	0.098*	0.124**	0.314
Λ_{33}	-0.251	1.052***	-0.053	-0.786*	-0.317
h_1	0.228***	0.142***	0.078***	0.081***	0.144***
h_2	0.337**	0.359***	0.319***	0.317***	0.416*
h_3	0.382***	0.180	0.526***	0.512***	0.467**
λ_1	0.007	-0.061	-0.751	-0.229	0.781
λ_2	0.000	2.003***	4.874***	4.267**	5.249***
λ_3	0.000	-0.003	3.527***	5.336***	0.000
b_1	-0.316**	-0.042	-0.002	-0.212	-0.142**
σ_f	0.006***	0.006***	0.006***	0.006***	0.006***
σ_e	0.146***	0.146***	0.141***	0.133***	0.129***

Notes: Significance levels are given by ***1%, **5% and *10%. Each column – after the parameter column – represents one set of parameters, obtained after the calibration process with the indicated years. The forecast that use each set of parameters is made for the following year after the calibration set of years.

As anticipated, the amount of data increases, starting from two years, to eleven years. For example, the column “2007-2013” contains the set of parameters obtained using seven years of data (from January 2007 to December 2013), and they are used to generate the forecast for 2014 (from January to December, both included).