



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Petroleum Inventory Level:
A Leading Indicator of Crude Oil Prices

Marcus Larsson

A Thesis Submitted for the Degree of Master of Science in Finance
Centre for Finance and Department of Economics, School of Business, Economics and
Law, Goteborg University
Email: guslarmadr@student.gu.se
Supervisor: Evert Carlsson

Abstract

This paper proposes a short-term forecasting model of West Texas Intermediate (WTI) crude oil spot prices using United States petroleum inventory levels and the spread between front-month futures prices and back-month futures prices. Applying the model between January 2010 to January 2018, I find that the model outperforms a naïve forecasting model with a convincing margin. The model is based on readily available data, which makes the model useful for those who are interested in forecasting future oil prices or wanting to understand historical price fluctuations.

Keywords. Petroleum Inventory, Crude oil price, Term Structure, Backwardation

J.E.L. classification. Q41, Q47, C22

I. ACKNOWLEDGMENT

Firstly, I would like to express my gratitude to my supervisor, Evert Carlsson, for his guidance and support throughout the process. I am very grateful to have been given the opportunity to pursue a topic I have great passion for. I would also like to thank the Graduate School of the University of Gothenburg and the Centre for Finance. Finally, I would like to give a shout-out to *The Organization Of Oil Trading Tweepers*, **#OOTT**, for providing oil news, events, and public data that rivals some of the best proprietary market analytics.

Marcus Larsson, Gothenburg, 28th May 2018

CONTENTS

I	Acknowledgment	3
II	Introduction	5
III	Theories	7
III-A	Relative Inventory	7
III-B	Non-linearity Between Relative Inventory and the Spot Price	8
III-C	Contango	9
III-D	The Theory of Normal Backwardation	10
III-E	The Theory of the Price of Storage	10
III-F	Properties of Spot and Futures Prices	11
IV	Model	13
IV-A	Model Specifications	13
IV-B	Causality and Endogeneity	14
V	Data	15
V-A	Petroleum Inventory	15
V-B	Term-Structure of the Futures Market	16
V-C	WTI and the Price Discovery Process	16
VI	Methodology	17
VI-A	Time series modeling	17
VI-B	Autoregressive Integrated Moving Average	17
VI-C	Econometric Specifications	18
VII	Empirical Results	19
VII-A	Regression Coefficients	19
VII-B	Hold-Out Cross-Validation	20
VII-C	K-fold Cross-Validation	23
VII-D	Leave-one-out Cross-Validation	23
VIII	Conclusion	24
	Appendix	25
A	Measurement of Seasonality	25
B	Price-Inventory Correlation	25
C	ARIMA Identification	26
D	Cross-Validation	27
E	Diebold–Mariano Tests	28

II. INTRODUCTION

In this paper I develop a dynamic forecasting model for the West Texas Intermediate (WTI) crude oil spot price. The objective of the thesis is to provide a simple and practical forecasting model which can easily be implemented with readily available data. ¹ I conduct four different forecasting models and study the performance of each model in an attempt to find the most accurate one. The first model uses the relative inventory level as suggested by [Ye et al. \(2002\)](#). The second model uses the spread between the front-month futures price and the back-month futures price. The third model is a combination of the first and the second model. The fourth model is a naïve forecasting model (random walk), which predicts that oil prices will stay at their current levels. In order to test the predictive power of the models, I employ pseudo out-of-sample forecasting. Consistent with [Ye et al. \(2002, 2005\)](#), I find evidence that the relative inventory model outperforms a naïve forecasting model. Moreover, I find evidence that the relative inventory model can be improved. This is shown as the third model, that is the combination of the first and second model, has the best accuracy out of the four models. The conclusion of my work is that the combined effect of the relative inventory level and the term-structure spread is a good market indicator of WTI spot price changes.

The use of the petroleum inventory level in explaining oil prices has been utilized in numerous previous forecasting models ([Ye et al., 2002, 2005](#); [Kaufmann et al., 2004](#); [Zamani, 2004](#); [Merino and Ortiz, 2005](#); [Byun et al., 2017](#)). One of the most prominent forecasting models is the relative inventory model, which normalizes the petroleum inventory level for seasonal movements and a general trend ([Ye et al., 2002](#)). From the early 1990's until the early 2000's, changes in the OECD relative inventory level demonstrated a strong negative correlation with changes in the WTI spot price ([Ye et al., 2002, 2005](#); [Zamani, 2004](#)). Following 2003, the relative inventory model started to fail, as it consistently under-predicted the oil price. [Merino and Ortiz \(2005\)](#) attributed the failure of the relative inventory model to increased participation of financial investors in the crude oil futures market. As a solution, [Merino and Ortiz \(2005\)](#) extended the relative inventory model to include money managers long positions in WTI futures-contracts and documented that such a specification could improve the model. This improvement has, however, later been disputed as the relationship falls apart following 2005 ([Byun et al., 2017](#)).

Due to the failure of the model, there has been little research related to the topic over the last decade. Nevertheless, the petroleum market has experienced several unique changes over the last years, which makes it interesting to revisit the relative inventory model. With the U.S. shale revolution, a paradigm shift has developed in the oil and gas industry. ² The surge in U.S oil production resulted in U.S. shale producers transforming into the world's new swing producers, leaving shale oil as the deciding factor for future prices ([Sachs, 2015](#)). ³ The exploration and development of shale oil requires lower investment of time and money, making shale oil production more responsive to price changes ([Bahgat, 2016](#)). Shale oil producer's ability to quickly adjust and

¹ The model deals with the fundamental relationship between inventories and prices and does not take into account geopolitical events, or financial crisis.

² The shale revolution refers to the use of horizontal drilling and hydraulic fracturing that enabled tremendous production growth of oil and natural gas in the U.S.

³ A swing producer is a supplier that has a large amount of spare capacity.

increase production levels resulted in a massive build of U.S. petroleum inventories, which contributed to the oil prices collapse in late 2014. As a result, I hypothesize that tracking the U.S. relative inventory level will help in explaining the historical price fluctuations over the last years. Going forward, I expect that the WTI spot price will largely be determined by shale producer's ability to ramp up production and whether future demand is strong enough to absorb these increased volumes. These elements will be directly captured in the U.S. relative inventory level and serve as the explanation for why I believe that the U.S. inventory level has grown to be the most imperative variable in its predictive power of WTI prices.

In similarity to [Merino and Ortiz \(2005\)](#), I try to improve the relative inventory model by examining the futures market. With the wide spread implementation of the futures market, the relationship between supply and demand has become more complex due to 'paper barrels' being created when producers and consumers actively hedge their risk. The relative inventory model focuses solely on the demand and supply in the physical market and fails to account for activity in the futures market. To account for market participants collective expectations and behavior outside the limits of the physical market, I incorporate the term-structure spread in the model. Research has shown that incorporating information on the relationship between the front-month futures prices and the back-month futures prices can help in predicting spot prices ([McCallum and Wu, 2005](#); [Boons and Prado, 2015](#); [Etienne and Mattos, 2016](#)). I hypothesize that the joint effect of the term-structure spread and the relative inventory level can be used to improve the traditional model as the two variables are directly connected to each other according to theory.

III. THEORIES

In this section, I present the theoretical perspective of the variables included in the forecasting model. In part A, I describe the relative inventory level, which suggests that oil prices are highly correlated to the petroleum inventory level. In part B, I discuss the non-linear relationship between the relative inventory level and the WTI spot price. In part C and D, I give a brief introduction to the concept of contango and backwardation. In part E, I describe the theory of the price of storage, which describes how the petroleum inventory level relates to the term-structure of the futures curve. In part F, I describe how the term-structure spread relates to the WTI spot price.

A. Relative Inventory

Intuitively, there should be a negative relationship between the petroleum inventory level and the oil price.⁴ The rationale is that a petroleum inventory change can be interpreted as the result of the imbalance between supply and demand, and should therefore affect the spot price in a negative manner.

$$\text{Inventory Change} = \underbrace{\text{Domestic Production} + \text{Imports}}_{\text{Supply}} - \underbrace{\text{Domestic Consumption} - \text{Exports}}_{\text{Demand}} \quad (1)$$

The relationship between the petroleum inventory level and the oil price is, however, not directly obvious, as the petroleum inventory level is affected by seasonal movements and a general trend that tends to mask the connection (See Appendix A). As originated by [Ye et al. \(2002\)](#), the relative inventory level is an indicator that normalizes the petroleum inventory level for seasonal variation in production, consumption and refinery utilization. The central concept is that if current inventories deviate from their normal level, according to inventory trend and seasonal swings, the market is in disequilibrium and prices should react accordingly.

The variables used for calculating the relative inventory are the following:

- IN_t The observed petroleum inventory level at the end of the week t , measured in millions of barrels.
- \widehat{IN}_t The normal inventory level at the end of week t , measured in millions of barrels.
- RIN_t The relative inventory level at the end of week t , derived as $IN_t - \widehat{IN}_t$ and measured in millions of barrels.
- D_k Monthly dummy variables measuring seasonality.

The observed inventories are de-seasonalized through the following regression.

$$IN_t = \sum_{k=1}^{12} \beta_k D_k + \epsilon_t \quad (2)$$

$$\widehat{IN}_t = \sum_{k=1}^{12} \beta_k D_k \quad (3)$$

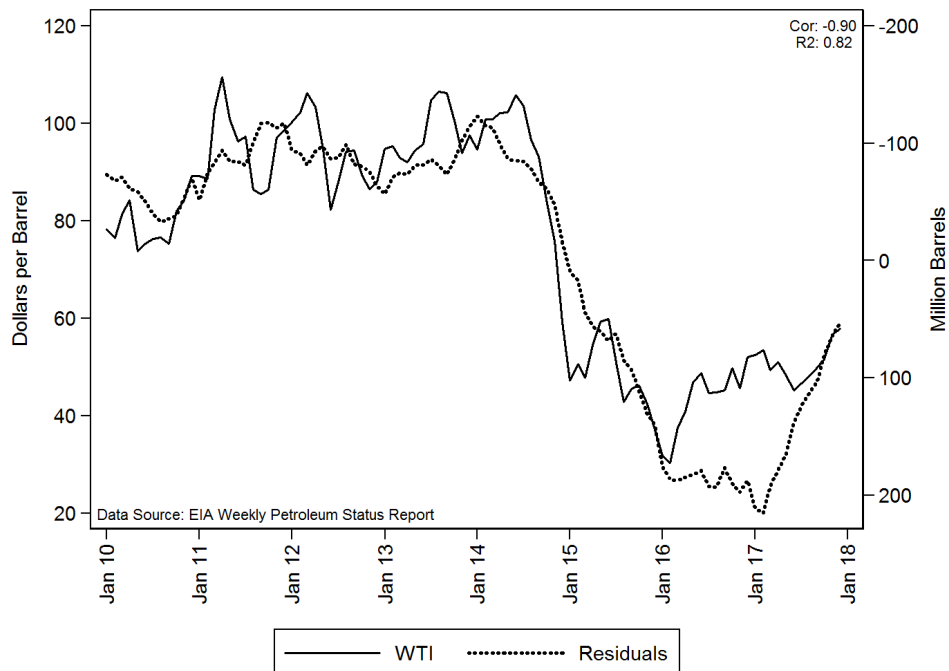
⁴ The petroleum inventory level is defined as the sum of crude oil and petroleum products. The reason for considering crude oil and petroleum products in aggregate is because crude oil inventories at refineries are essentially exchangeable with product inventories in a monthly time frame.

Once the dummy parameters are estimated, one can extract the seasonal influence of the inventory level, where the resulting residuals will represent the relative inventory level.

$$RIN_t = IN_t - \widehat{IN}_t = \epsilon_t \quad (4)$$

Figure 1 below demonstrates the negative relationship between the WTI spot price and the relative inventory level. The horizontal axis shows the time period between January 2010 to January 2018. On the left vertical axis, the WTI spot price is shown in U.S. dollars per barrel and on the right vertical axis, the U.S. relative inventory level is shown in million barrels. The relative inventory level is shown in inverse for a clearer visualization of the relationship. The correlation over the eight year period is as high as -0.9.

Figure 1: The WTI Spot Price and the Relative Inventory Level



B. Non-linearity Between Relative Inventory and the Spot Price

A minimum operating level of petroleum inventory is essential to keep the North American supply system operating; pipeline systems need a cushion level of inventory to keep the system running, road tankers and railcars need fuel to link the production sites, and terminals and refineries need a base level to operate. Due to the economy requiring a minimum operating level of petroleum inventories, the relationship between the inventory level and the spot price is intrinsically non-linear (Ye et al., 2006). That is, if the inventory level were to approach its minimum operating level, prices should in theory react in a non-linear fashion to compensate for the risk associated with low inventory levels.

Petroleum inventories can also be viewed as being limited due to infrastructure constraints. Once the inventory surplus breaches logistical and spare storage capacity, oil prices should converge to marginal cost and force producers

to stop producing. U.S. shale wells can have a variable cost below 15 dollars a barrel, resulting in that their owners will keep on producing even if spot prices are below the producers average cost (Kleinberg et al., 2016).

Figure 2: Non-Linear Relationship Between the WTI Spot Price and the Relative Inventory Level

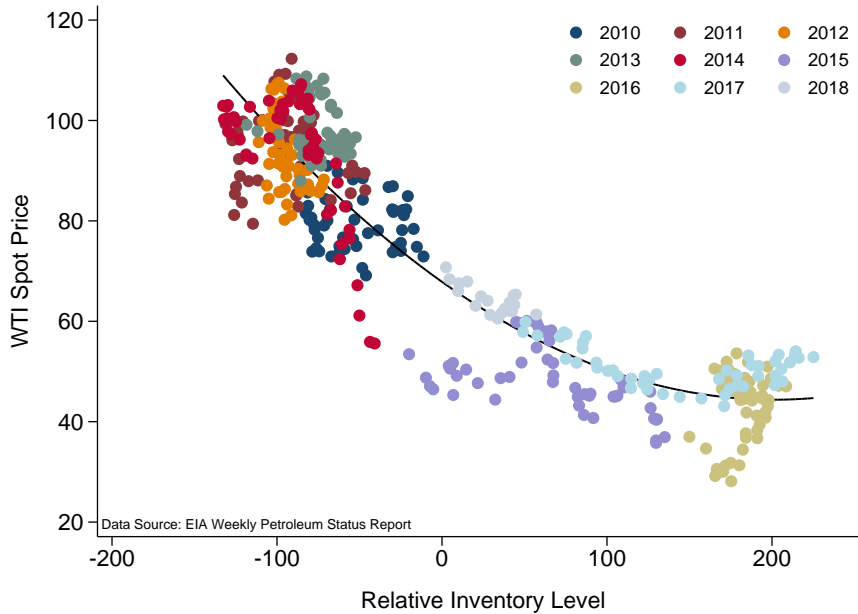


Figure 2 above, demonstrate the non-linear relationship between the WTI spot price and the relative inventory level.⁵ On the vertical axis, the WTI spot price is shown in U.S. dollars per barrel and on the horizontal axis, the U.S. relative inventory level is shown in billion barrels. The upper and lower bound of petroleum inventories - created by infrastructure constraints and the minimum operating level - contributes with a steeper curvature at the extreme points. In essence, the market recognize whether the prevailing storage level is sustainable or not and adjusts the WTI price accordingly. The price is thus the mechanism the market employs to secure adequate supply and to discourage the maintenance of surplus supply (Bodell, 2009).

C. Contango

When futures prices are trading at a premium to spot prices, the market is said to be in contango. Contango emerges with the view that there is a surplus of oil on the market, and that current demand is weak, which reduces price at the front-end of the term-structure. A market in slight contango is often considered as the default state for a storable commodity. Intuitively, distant-month futures should be priced slightly higher than near-month futures in order to justify the cost of storage and interest (Bouchouev, 2012). Imagine a scenario of strong contango, where the distant-month future is priced significantly higher, than the near-month future, exceeding the costs of storage and the interest rate. Such a scenario creates incentives for market participants with physical storage to arbitrage by buying the near-month future and selling the far-month future. The market participant captures the term-structure spread by taking delivery of the oil in the near-month, storing the oil, and then providing delivery in the far-month. As more market participants with storage take advantage of this opportunity, they will flatten the term-structure until contango equals the equivalent of the cost of storage and interest. Without storage capacity constraints, elevated

⁵ The fitted curve is derived by regressing the WTI spot price on the relative inventory level and the relative inventory level squared.

levels of contango will not be sustainable, while a modest contango can be considered ordinary (Alquist and Kilian, 2010).

D. The Theory of Normal Backwardation

When futures prices are trading at a discount to spot prices, the market is said to be in backwardation. Backwardation characterizes a market where oil is in shortage, and where current demand is strong. The presence of backwardation used to be viewed as a state of abnormality, as it appeared irrational for market participants to be willing to sell the far-month future despite the cost of storage and the interest rate. Keynes (1930) and Hicks (1939) explained this phenomenon by developing the theory of normal backwardation. Normal backwardation argues that the futures market is setup for hedging purposes; resulting in commodity producers selling futures at a discount to the spot price in order to create incentives for external capital to take on the producer's risk of falling prices. The conclusion of the theory is that physical actors are generally net short while speculators are generally net long. The theory further assumes that the cash price is determined by the supply and demand in the physical market, while the futures price is equal to the expected cash price minus the risk premium hedgers are willing to pay. Keynes would thus argue that the futures price is a downward biased estimate of the future cash price.

E. The Theory of the Price of Storage

The traditional theory of storage, as developed by Working (1949), is very similar to the theory of normal backwardation, but slightly more comprehensive. The theory of storage believes that cash and futures prices aren't autonomous, with spot and futures prices being determined together rather than in isolation. More specifically, Working said that inter-temporal prices reflect supply and demand for inventories. In equation 5, the no-arbitrage condition, as derived from the theory of the price of storage, is expressed as:

$$F_{t,T} = S_t(1 + r_{t,T}) + W_{t,T} - C_{t,T} \quad (5)$$

where $F_{t,T}$ is the futures price at time t for delivery at T , S_t is the spot price at time t , $r_{t,T}$ is the risk-free rate, $W_{t,T}$ is the warehouse cost and $C_{t,T}$ is the convenience yield. The convenience yield is defined as the benefit of holding the underlying physical commodity, rather than the derivative product. The convenience yield represents the benefit of having the physical product immediately at hand in case of an emergency like a temporary shortage. The marginal convenience yield, i.e., the benefit of holding the next unit of inventory, falls as inventories build and rises when inventories draw. The rationale is that one extra unit of inventory is more valuable if inventories are scarce and less valuable if inventories are in abundance (Working, 1949; Kaldor, 1939; Brennan, 1958; Telser, 1958).

As an implication of the theory of storage, the term-structure of futures prices is determined by the interest rate, the storage cost and the convenience yield. Assuming that the interest rate and the storage cost are constant, the theory implies that the term-structure is shaped by the inventory level over time. In more specific, the futures market

will be in backwardation when the petroleum inventory level is low and in contango when the petroleum inventory level is high. This is demonstrated in equation 6.

$$S_t - F_{t,T} = C_{t,T} - W - S_t r \quad (6)$$

F. Properties of Spot and Futures Prices

A large literature has examined futures prices ability to forecast spot prices. Coppola (2008), Reichsfeld et al. (2011), Reeve and Vigfusson (2011), Chinn and Coibion (2014) and Fama and French (2016), among others, find evidence suggesting that futures prices have predictive power over spot prices. Whereas equally large amount of research like Bopp and Lady (1991), Moosa and Al-Loughani (1994), Chernenko et al. (2004), Alquist and Kilian (2010) and Alquist et al. (2013) find no compelling evidence supporting this thesis. Futures prices of non-storable commodities embody only market expectations of future supply and demand conditions. For these commodities, there exists no arbitrage opportunities, resulting in futures prices representing an unbiased predictor of the future spot price. For storable commodities, like oil, futures prices are generated by arbitrage, resulting in futures prices not necessarily being a good predictor of the future spot price (Daniel et al., 1999). Rather than considering the futures price of WTI in isolation, research has shown that incorporating information on the relationship between front-month futures prices and back-month futures prices can help in predicting oil prices (McCallum and Wu, 2005; Boons and Prado, 2015; Etienne and Mattos, 2016; Jin, 2017). The rational is that there is a forward looking element embedded in the term-structure which can be exploited to improve forecast accuracy.

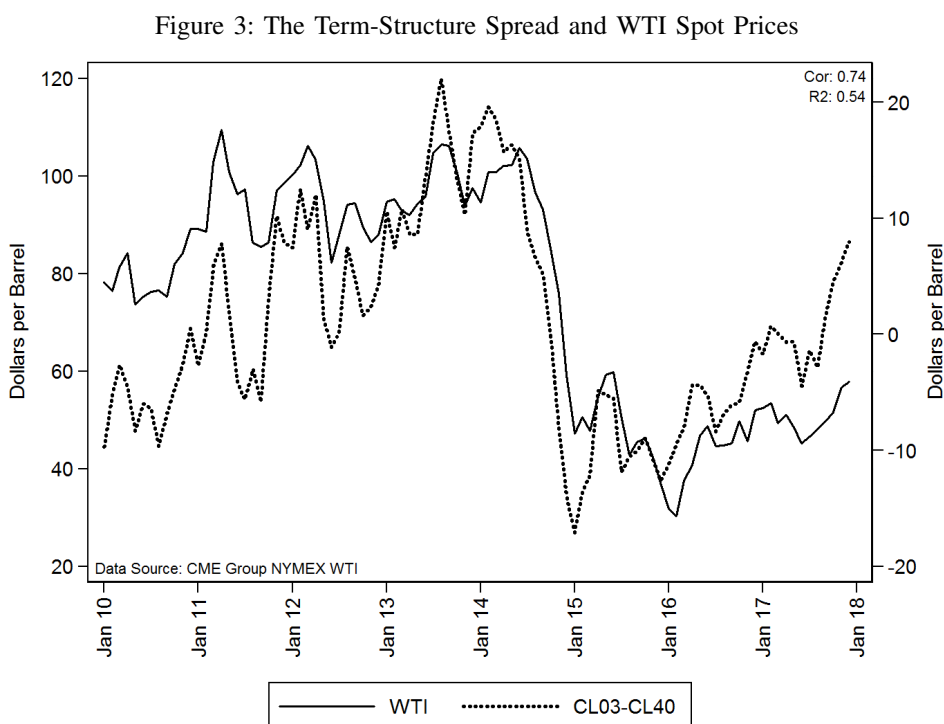


Figure 3 above, demonstrate the relationship between the term-structure spread and the WTI spot price. On the left vertical axis, the WTI spot price is shown in U.S. dollars per barrel and on the right vertical axis, the

term-structure spread is shown in U.S. dollars per barrel. The term-structure spread is defined as the spread between the front-month futures price and the back-month futures price. When the term-structure spread is positive, the market is in backwardation and when the term-structure spread is negative, the market is in contango. The abbreviation for WTI crude oil futures are denoted CL, where CL03 represent the futures contract prompt in 3 months and CL40 represent the futures contract prompt in 40 months. The data suggests that rising prices should be coextensive with an increasing term-structure spread and that falling prices should be coextensive with a decreasing term-structure spread. In other words, backwardation is considered bullish for oil prices while contango is considered bearish for oil prices.

A backwardated market is considered bullish for oil prices as:

- it's a sign of tightening supplies and strong demand (low inventory level).
- it punishes those who store oil, which will result in lower inventory levels in the future.
- it contributes with a positive roll yield, i.e. the market pays you to be long.⁶
- it limits shale producers ability to hedge future cash flows.

A market in contango is considered bearish for oil prices due to the exact opposite reasons.

⁶ The roll yield is the yield captured given that a long position in a futures contract converges to the spot price. (de Groot et al., 2014).

IV. MODEL

In this section, I derive four alternative models to predict the oil price. In subsection A, I determine which independent variables should be included or excluded in the different models. In subsection B, I identify the causal relationship to validate the chosen variables and discuss potential endogeneity problems of the models.

A. Model Specifications

I formulate four different forecasting models in an attempt to find the most accurate one. The first model, which is called the relative inventory model, uses the relative inventory level. The model rest on the notion that the price of oil is determined by the equilibrium between supply and demand in the physical market. In equation 7, the relative inventory model is expressed as:

$$WTI_t = \alpha + \beta_1 RIN_{t-1} + \epsilon_t \quad (7)$$

where WTI_t is the average weekly WTI spot price at time t and RIN_{t-1} is the relative inventory level a period before.

The second model, which is called the term-structure spread model, uses the spread between the 3:rd-month WTI futures contract less the 40:th-month WTI futures contract. The model builds on the assumption that the futures market contains informational content regarding market participants expectations and behavior which can help in forecasting WTI prices. In equation 8, the term-structure spread model is expressed as:

$$WTI_t = \alpha + \beta_1 TS_{t-1} + \epsilon_t \quad (8)$$

where WTI_t is the average weekly WTI spot price at time t and TS_{t-1} is the term-structure spread a period before.

The third model is a combination of the first and the second model. A squared variable of the relative inventory is added to capture the non-linear dynamic between oil prices and petroleum inventories. In equation 9, the combined model is expressed as:

$$WTI_t = \alpha + \beta_1 RIN_{t-1} + \beta_2 RIN_{t-1}^2 + \beta_3 TS_{t-1} + \epsilon_t \quad (9)$$

where RIN_{t-1} is the relative inventory a period before, RIN_{t-1}^2 is the squared relative inventory a period before and TS_{t-1} is the term-structure spread a period before.

The fourth model is a naive forecasting model. The model rest on the notion that oil prices follows a random walk. The naïve forecasting model is used as a benchmark to evaluate the forecasting performance of the other

models. In equation 10, the naive forecasting model is expressed as:

$$WTI_t = WTI_{t-1} \quad (10)$$

where WTI_t is the average weekly WTI spot price at time t and WTI_{t-1} is the average weekly WTI spot price a period before.

B. Causality and Endogeneity

Following [Merino and Ortiz \(2005\)](#), I identify the causal relationship between the chosen variables to ensure validity of the models. In order to examine the causal relationship, I employ the Toda and Yamamoto Non-Causality Test. The [Toda and Yamamoto \(1995\)](#) version of the Granger non-causality test allows for testing causality between parameters regardless of whether the processes may be integrated or cointegrated of an arbitrary order. Rather than testing the first differences, as is often the case with Granger causality, the approach fits an autoregressive model in the levels of the variables and thereby minimizes the risk of wrongly identifying the order of integration.

Table I: Toda-Yamamoto Non-Causality Test

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Chi-square</i>	<i>P-value</i>	<i>Causality</i>	<i>Cointegration</i>
Relative Inventory	WTI	21.54	0.00	Yes	Rank 1
WTI	Relative Inventory	4.73	0.32	No	Rank 1
Term-Structure Spread	WTI	39.06	0.00	Yes	Rank 1
WTI	Term-Structure Spread	6.60	0.04	Yes	Rank 1
Term-Strucure Spread	Relative Inventory	2.56	0.46	No	Rank 0
Relative Inventory	Term-Structure Spread	11.08	0.01	Yes	Rank 0

*The number of optimal lags is determined by using the FPE, the AIC, the HQIC and the SBIC. The presence of serial correlation in the residuals are tested for using the Lagrange multiplier test. The order of cointegration rank is tested by using the Johansen Cointegration test.

Table I presents the results from the causality test. The conclusions from the causality tests at five percent significance level are the following:

- The relative inventory level predicts the WTI spot price.
- The term-structure spread predicts the WTI spot price.
- The WTI spot price predicts the term-structure spread.
- The relative inventory level predicts the term-structure spread.

For causal inference, the major goal is to get unbiased estimates of the regression coefficients. As a result, endogeneity problems represents a major concern when examining causal inference ([Berman et al., 2011](#); [Antonakis et al., 2014](#)). With predictive modeling, however, endogeneity is much less of an issue. In predictive modeling the intent is not to get the optimal estimates of the true coefficients. Rather, the ambition is to get the most accurate prediction based of the variables which are available. Endogeneity is a concern only insofar as one might be able to improve the prediction by removing the endogeneity ([Shmueli, 2010](#); [Allison, 2014](#)). Hence, the bidirectional relationship between the term-structure spread and the WTI spot price (as suggested by the causality test) might not represent a big problem.

V. DATA

A database covering petroleum inventory is a prerequisite for this work. The International Energy Agency (IEA) and the U.S. Energy Information Administration (EIA) provides the most comprehensive data in terms of quality, nations covered, consistency of reporting and detailed used. I will use the *EIA Weekly Petroleum Status Report*, as the data is readily available online for anyone to reach. The EIA data set is published on a weekly basis, and covers U.S. petroleum inventories by Petroleum Administration for Defense District (PAD District). I chose to limit the data to the United States, as I hypothesize that the U.S. inventory level has grown to be the most imperative region in its predictive power of WTI prices.

A. Petroleum Inventory

The Petroleum inventory level represent the amount of inventory of crude oil and petroleum products held in inventory for future use. Inventories are accounted for on a national territory basis, within a country's geographical region and irrespective of ownership. Granted that the inventories are held on the national territory, it does not matter whether the inventories are held onshore, offshore, at refineries or in pipelines. The inventory level is measured by the *EIA-813 Monthly Crude Oil Report*, which requires companies that carry or store 1,000 barrels or more to submit information regarding all domestic and foreign inventories held in custody and in transit thereto.⁷ The petroleum inventory consists of crude oil, which represent the liquid that is extracted from the geological formation, and petroleum products, which are produced from the processing of crude oil. Petroleum inventories can further be divided into commercial inventories, which represent petroleum inventories held for commercial purposes by U.S. firms, and strategic inventories, which represent petroleum inventories maintained by the Federal Government.

Consistent with previous research, I find that the price correlation is at its strongest when considering both crude oil and petroleum products together rather than one in isolation (See Appendix B). This makes intuitive sense, as abnormal behavior in crude oil or petroleum products could be missed if fixating on one in isolation. For this reason, I decide to use the sum of crude oil and petroleum products when calculating the relative inventory level. While most of the earlier literature uses the sum of crude oil and petroleum products when calculating the relative inventory level, there exists differences between the choice of including the strategic petroleum reserve (SPR) or not. For instance, [Merino and Ortiz \(2005\)](#) use the commercial inventory level - with the reason that the SPR could generate a miss-specification of the equilibrium given its strategic nature - while [Ye et al. \(2003\)](#) carry out analysis of the total inventory level due to several countries reclassifying inventories between the governmental and the commercial category. Even though I recognize that government inventories are generally for strategic reasons and are not always determined by market forces, I choose to include the SPR while calculating the relative inventory level. Over the years between 2010 to 2018, the SPR has been falling with around 60 million barrels; with approximately half being released between July to September in 2011 to offset the ongoing supply distributions in Libya, and the rest

⁷ The Weekly Petroleum Status Report is partly based on projected data, while the EIA-813 Monthly Crude Oil Report uses the correct numbers, but are lagged two months behind. The weekly data are revised to match the monthly when the correct data is available.

being released continuously between 2017 and 2018. In my view, it would be incorrect to ignore the SPR over the selected time period, due to U.S. demand absorbing these releases.

B. Term-Structure of the Futures Market

The abbreviation for WTI crude oil futures are denoted CL#. CL has its contract list sorted by month, where CL01 represent the 1st generic contract, i.e. the contract that is closest to expiration at any given point in time, and CL02 represent the contract following the CL01 future, and so on. I define the term-structure spread as the difference between CL03 and CL40, with the reasoning that the CL03 contract is not directly included in the price discovery process and the CL40 being a valid estimate for the long run futures price while still being liquid enough.

C. WTI and the Price Discovery Process

For crude oil spot prices, I use West Texas Intermediate (WTI), which together with Brent are considered the world's benchmark for crude oil price. WTI is the crude oil specified for delivery to Cushing, Oklahoma under the New York Mercantile Exchange (NYMEX) futures contract, which makes WTI the proper benchmark to compare with U.S. petroleum inventories.⁸ The NYMEX WTI is the most liquid commodity future in the world and underpins an important role in the price discovery process of the WTI spot price (Geman and Kharoubi, 2008). As market participants prefer the most accurate value for their crude oil, the WTI spot price is linked to the highly liquid NYMEX futures contract, which is easy to hedge with and hard to manipulate. The daily settlement price or close for the adjacent NYMEX crude futures contract serve as the benchmark price for the WTI crude oil spot market. The most common formula for determining the physical pricing element in the NYMEX future is the NYMEX calendar month average (CMA). The basic CMA is simply the average of the NYMEX closing price during the month when the physical crude is delivered.

⁸ In Appendix B, I show that the correlation between the relative inventory level and the crude oil price is as strongest when matching U.S inventories with WTI spot prices and OECD inventories with Brent spot prices.

VI. METHODOLOGY

In this section, I start by giving a short background and description of the ARIMA model and motivate why I use it as my forecasting tool. The econometric specifications are then listed under subsection C.

A. Time series modeling

The classical linear regression model builds on the assumption of independent and identically distributed observations. For cross-sectional data, independence between observations is automatically fulfilled when random sampling is used. For time series, one typically cannot assume that the samples which are taken throughout time are independent of one another. Time series tend to contain a high degree of auto-correlation, which is particularly the case if the sampling interval is small, such as a week or a month. Furthermore, time series data tend to be non-stationary in levels, which violates the requirement of identically distributed observation. By employing a classic linear regression model for time series data, the risk of producing a spurious model is high because sufficient care is not taken during formulation of the auto-correlation structure and non-stationarity of the data (Granger and Newbold, 1974) (Adhikari and Agrawal, 2013). Three major consequences of auto-correlation and non-stationarity in the data in regression analysis are the following:

- Estimates of the regression coefficients are biased.
- Forecasts based on the regression equations are sub-optimal.
- The usual significance tests on the coefficients are invalid.

In order to avoid these problems, the Autoregressive Integrated Moving Average model was selected for the study.

B. Autoregressive Integrated Moving Average

One of the most general and commonly used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model includes an explicit statistical model for the irregular component of a time series, that allows for non-zero auto-correlations in the irregular component. The ARIMA model further allows for an initial differencing step, which eliminate the non-stationarity in the data. The order of an ARIMA is usually denoted by the notation ARIMA(p,d,q), where

- p is the order of the autoregressive part
- d is the order of differencing
- q is the order of the moving-average process

Mathematically the ARIMA model is written as

$$(1 - \beta)^d Y_t + \mu + \frac{\theta(\beta)}{\phi(\beta)} \alpha_t \quad (11)$$

where,

- t indexes time
 μ is the mean term
 β is the backshift operator; that is, $\beta X_t = X_{t-1}$
 $\theta(\beta)$ is the autoregressive operator, represented as a polynomial in the back shift operator
 $\phi(\beta)$ is the moving-average operator, represented as a polynomial in the back shift operator
 α_t is the independent disturbance, also called the random error

The ARIMA (1,1,0) model is chosen for this study. The motivation behind the identification of the ARIMA order, i.e. the identification of (p,d,q), can be found in Appendix C.

C. Econometric Specifications

The econometric specification of the four models are listed below. The variables used are as following:

- WTI_t The average weekly WTI spot price, measured in dollars per barrel.
 RIN_t The relative inventory level at the end of week t, measured in millions of barrels.
 RIN^2_t The relative inventory level squared.
 TS_t The term-structure spread at the end of week t.

The relative inventory model is specified in equation 12.

$$\Delta WTI_t = \Delta \sum_{i=1}^2 \beta_i RIN_{t-i} + \Delta c WTI_{t-1} + \Delta \mu_t \quad (12)$$

The term-structure spread model is specified in equation 13.

$$\Delta WTI_t = \Delta \sum_{j=1}^2 \beta_j TS_{t-j} + \Delta c WTI_{t-1} + \Delta \mu_t \quad (13)$$

The combined model is specified in equation 14.

$$\Delta WTI_t = \Delta \sum_{i=1}^2 \beta_i RIN_{t-i} + \Delta \sum_{j=1}^2 c_j TS_{t-j} + \Delta \sum_{k=1}^2 d_k RIN^2_{t-k} + \Delta e WTI_{t-1} + \Delta \mu_t \quad (14)$$

The naive forecasting model is specified in equation 15.

$$WTI_t = WTI_{t-1} \quad (15)$$

VII. EMPIRICAL RESULTS

In order to test the predictive power of the models, I employ pseudo out-of-sample forecasting. I compare the four models to the actual WTI spot price by using three different cross-validation techniques. I also compare the first three models with the naïve forecasting model (random walk), which can be surprisingly difficult to beat in practice. I find that the three first models outperform the naïve forecasting model with statistical significance. This result suggests that the petroleum inventory level has been a leading indicator for WTI spot prices for the last decade. Out of the four models, I find that the combined model has the best performance. This result suggests that the combined effect of the relative inventory level and the term-structure spread can be used to improve forecast accuracy. In subsections A, I present the regression coefficients of the different models. In subsection B, C and D, the forecasting results for the various evaluation methods are shown.

A. Regression Coefficients

Table II: The Relative Inventory Model

<i>Variable</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>t-statistic</i>	<i>P-value</i>
RIN_{t-1}	(0.036)	0.014	(2.52)	0.012
RIN_{t-2}	(0.031)	0.014	(2.16)	0.031
WTI_{t-1}	0.243	0.053	4.58	0.000
Adj. R^2 (diff)	Adj. R^2 (lev)	Durbin-h	AIC	SBIC
0.086	0.990	0.242	1877	1893

*The adjusted R-square is shown by running the regression in first difference and in levels. The Durbin h statistics is used to examine autocorrelation of the errors for models with dynamic specifications (Durbin, 1970). The Akaike's information criterion (AIC) and the Schwarz's Bayesian information criterion (SBIC) is used to evaluate the goodness of fit of the model. A lower number of the AIC and SBIC is preferable as it suggest a well specified model (Akaike, 1974; Schwarz et al., 1978).

The regression coefficients from the relative inventory model, demonstrated in Table II, show that the relative inventory level is negatively related to oil prices, with both the lag one and lagg two coefficients of RIN_t being negative and statistical significant. The result suggests that a one unit change in the relative inventory level results on average in a -0.09 dollars change in the spot price.⁹ This is equivalent to a 0.63 dollar price change for an one million barrel-per-day change in a week.

Table III: The Term-Structure Spread Model

<i>Variable</i>	<i>Coefficient</i>	<i>Standard error</i>	<i>t-statistic</i>	<i>P-value</i>
TS_{t-1}	0.639	0.083	7.73	0.000
TS_{t-2}	0.102	0.074	1.37	0.170
WTI_{t-1}	(0.024)	0.0069	(0.36)	0.722
Adj. R^2 (diff)	Adj. R^2 (lev)	Durbin-h	AIC	SBIC
0.186	0.999	0.068	1827	1839

The regression coefficients from the term-structure spread model, demonstrated in Table III, show that the term-structure spread is positively related to oil prices, with both the lag one and lagg two coefficients of TS_t being positive. The result suggests that a one unit change in the term-structure spread results on average in a 0.76 dollar

⁹ The price effect is calculated from the model coefficients by $\frac{\text{Sum of RIN coefficients}}{1-\text{coefficient of } WTI_{t-1}}$, or $\frac{-0.067}{1-0.024}$.

change in the spot price. The autoregressive tendencies of the term-structure model appears significantly lower than of the relative inventory model, with the WTI_{t-1} coefficient being statistical insignificant for the term-structure spread model.

Table IV: The Combined Model

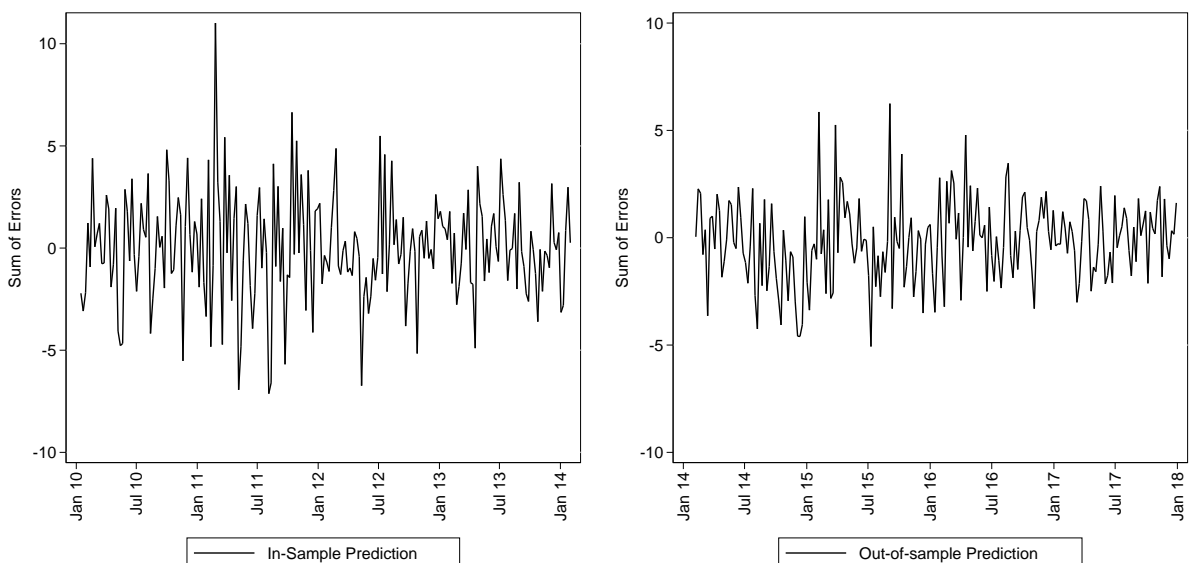
Variable	Coefficient	Standard error	t-statistic	P-value
RIN_{t-1}	(0.038)	0.014	(2.71)	0.007
RIN_{t-2}	(0.018)	0.013	(1.36)	0.176
RIN^2_{t-1}	0.00004	0.00005	0.78	0.435
RIN^2_{t-2}	0.00001	0.00005	0.12	0.902
TS_{t-1}	0.625	0.083	7.51	0.000
TS_{t-2}	0.089	0.075	1.19	0.235
WTI_{t-1}	(0.031)	0.068	(0.46)	0.647
Adj. R^2 (diff)	Adj. R^2 (lev)	Durbin-h	AIC	SBIC
0.198	0.999	0.004	1824	1853

The results of the combined model, shown in Table IV, is consistent with the previous results, which suggest a negative relationship between the relative inventory level and oil prices and a positive relationship between the term-structure spread and oil prices. The coefficient of RIN^2_t is, as expected, positive (see figure 2, page 7). The AIC suggest that the combined model is the preferred model, while the SBIC suggest that the term-structure spread is the preferred model.¹⁰

B. Hold-Out Cross-Validation

In the hold-out method, I divide the sample approximately in half, with the first period being defined as in-sample and the later period being defined as out-of-sample. This forces the model to outperform without incorporating any data from the oil crash into the training period. See Appendix D for a detailed explanation of the different cross validation technique used in this study.

Figure 4: Prediction Errors for the Relative Inventory Model



¹⁰ The AIC and SBIC measures the trade of between the numbers of parameters used in the model and the goodness of fit of the model. The measurements penalize overly complex models and helps to prevent overfitting. The AIC is generally viewed to be better-suited for model selection intended for prediction purposes (Sober, 2002; Shmueli, 2010)

Figure 4 illustrates the forecast errors for the relative inventory model for both in-sample and out-of-sample prediction. The y-axis measures the difference between the WTI spot price and the predicted values. Ideally, the sum of the errors would be zero, which is obviously not possible in practice. Furthermore, we want over-predictions and under-predictions to reflect the distribution of a random process, as its signals that the model is unaffected by the presence of bias. The relative inventory model appears to be well behaved, as the model under- and overpredicts with equal frequency. The worst performance of the relative inventory model is in February 2011, when the model under-predicted spot prices by 11 dollars per barrel. The under-prediction is attributed to the first Libyan Civil war. The crisis in Libya raised concern that the turmoil in the middle east could spread to other producing countries like Saudi Arabia. This led to an 11 dollar per barrel rally in the last week of February 2011.

Figure 5: Prediction Errors for the Term-Structure Model

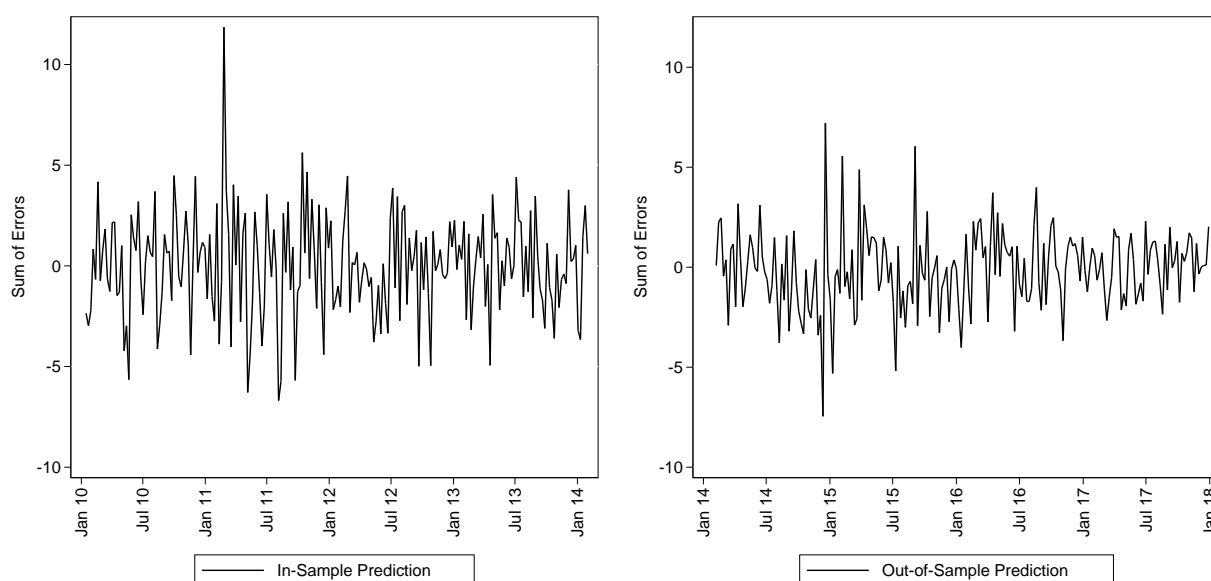


Figure 5 illustrates the forecast errors for the term-structure spread model. The y-axis measures the difference between the actual WTI spot price and the predicted values. When comparing the results of the term-structure model and the relative inventory model, December of 2014 stands out. For the last two weeks of 2014, oil prices declined by around 11 dollars per barrel. The decline was attributed to Saudi Oil Minister Ali al-Naimi reiterating Saudi's inaction to cut production. The term-structure spread model appears to perform worse than the relative inventory model for this period.

Figure 6 illustrates the forecast errors for the combined model. The y-axis measures the difference between the WTI spot price and the predicted values. The combined model appears to behave very similar to the term-structure spread model even if not identical.

Figure 6: Prediction Errors for the Combined Model

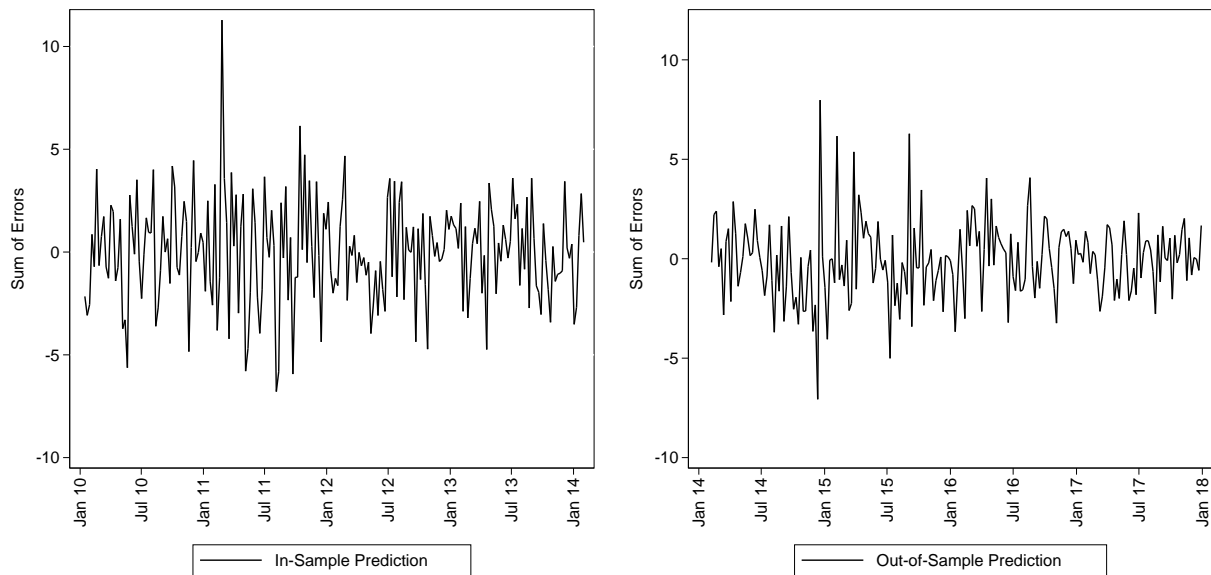


Table V: FCSTATS: C.F Baum

	<i>Out-of-Sample</i>				<i>In-Sample</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
RMSE	1.94	2.03	2.00	2.05	2.64	2.54	2.51	2.73
MAE	1.50	1.51	1.50	1.61	2.01	1.99	1.96	2.09
MAPE	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02
Theil's U	0.96	0.95	0.98	1	0.97	0.93	0.92	1

*Model 1 represents the relative inventory model, model 2 represents the term-structure spread model, model 3 represents the combined model and model 4 represents the naive forecasting model.

Table V lists several measurements of the different forecasting errors for the four tested models. The measurements include root mean squared error (RMSE), mean absolute error (MAE), mean absolute percent error (MAPE) and Theil's U (the uncertainty coefficient) (Baum, 2017). Model 1 represents the relative inventory model, model 2 represents the term-structure spread model, model 3 represents the combined model and model 4 represents the naive forecasting model. The table shows that the relative inventory model performs best out-of-sample while the combined model performs best in-sample.

Given the appeal of a formal statistical procedure for forecast comparison, I employ the Diebold–Mariano Tests (Baum, 2011). The test measures the statistical significance of divergence between the forecast models by testing the null hypothesis of no difference in accuracy. Hence, the test explains if the out-performance is statistically significant or due to randomness (see Appendix E for more information). Table VI compares the performance of the three models against the naive forecasting model. As with the previous table, model 1 represent the relative inventory model, model 2 represent the term-structure spread model, model 3 represent the combined model and model 4 represent the naive forecasting model. The table demonstrates that the three first models outperform the naive forecasting model with statistical significance.

Table VI: Diebold-Mariano Comparison

<i>Competing forecasts</i>	<i>Out-of-sample</i>	<i>In-Sample</i>
MAE Model 1	1.50	2.01
MAE Model 4	1.61	2.08
MAE Difference	(0.11)	(0.08)
p-value	0.00	0.00
MAE Model 2	1.51	1.99
MAE Model 4	1.61	2.10
MAE Difference	(0.10)	(0.10)
p-value	0.01	0.06
MAE Model 3	1.50	1.96
MAE Model 4	1.61	2.08
MAE Difference	(0.11)	(0.12)
p-value	0.01	0.00

*Model 1 represents the relative inventory model, model 2 represents the term-structure spread model, model 3 represents the combined model and model 4 represents the naive forecasting model.

C. K-fold Cross-Validation

In the K-fold cross validation, I partition the data set into five subsamples of equal size. I then pick four subsamples as the training set, and one subsample as the evaluation set. This procedure is run five times, using each subsample once as the evaluation set (See Appendix D).

Table VII: K-fold CV

	<i>RMSE</i>				<i>MAE</i>			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Estimate 1	2.24	2.45	2.23	2.48	1.74	1.49	1.75	2.24
Estimate 2	2.47	2.27	2.41	2.42	1.53	1.84	1.89	1.81
Estimate 3	2.39	2.27	2.29	2.31	2.11	2.09	1.85	1.55
Estimate 4	2.32	2.52	2.32	2.33	1.79	1.62	1.59	1.71
Estimate 5	2.42	2.19	2.14	2.49	1.92	1.89	1.77	1.92
Average	2.37	2.34	2.28	2.41	1.82	1.79	1.77	1.85

*Model 1 represents the relative inventory model, model 2 represents the term-structure spread model, model 3 represents the combined model and model 4 represents the naive forecasting model.

Table VII lists the root mean squared errors (RMSE) and the mean absolute errors (MAE) for the K-fold cross validation (Daniels, 2012). The table shows that the combined model performs best.

D. Leave-one-out Cross-Validation

The Leave-one-out cross validation is similar to the k-fold cross validation. The leave-one-out cross validation, however, uses a substantially higher number of subsamples, which produces a higher accuracy.

Table VIII: LOOCV

	Model 1	Model 2	Model 3	Model 4
RMSE	2.33	2.32	2.29	2.36
MAE	1.82	1.77	1.74	1.79

*Model 1 represents the relative inventory model, model 2 represents the term-structure spread model, model 3 represents the combined model and model 4 represents the naive forecasting model.

Table VIII lists the root mean squared errors (RMSE) and the mean absolute errors (MAE) for the leave-one-out cross validation (Barron, 2014). The table shows that the combined model performs best.

VIII. CONCLUSION

This paper presents and evaluates a dynamic forecasting model of WTI spot prices using U.S. petroleum inventories and the spread between front-month futures prices and back-month futures prices. The result suggests that the combined effect of the relative inventory level and the term-structure spread is a good market indicator of WTI spot price changes. The model uses historical data to predict future oil prices. For future research, I suggest that possible improvements could be made by forecasting the petroleum inventory level rather than using historical values. The change in the petroleum inventory level is derived by the following equation.

$$\textit{Inventory Change} = \textit{Domestic Production} + \textit{Imports} - \textit{Domestic Consumption} - \textit{Exports}$$

Domestic production can be forecasted by estimating the number of producing oil wells and their respective production rate. Imports and exports can be forecasted by tracking vessel flows, volumes in an out of ports and the spread between WTI and Brent prices. Domestic consumption can be forecasted by using data for per capita economic activity.

APPENDIX

A. *Measurement of Seasonality*

Table IX: Measurement of Seasonality

	<i>Crude oil</i>		<i>Petroleum Products</i>		<i>Crude & Products</i>	
	<i>Coef.</i>	<i>Seasonality</i>	<i>Coef.</i>	<i>Seasonality</i>	<i>Coef.</i>	<i>Seasonality</i>
Jan	1067	(19)	759	(3)	1826	(22)
Feb	1081	(5)	748	(13)	1829	(18)
Mar	1095	9	732	(30)	1827	(21)
Apr	1105	19	731	(30)	1836	(11)
May	1105	18	740	(21)	1845	(3)
June	1103	17	760	(1)	1863	16
July	1091	5	778	17	1869	22
Aug	1075	(10)	785	23	1860	13
Sept	1077	(9)	797	36	1874	27
Oct	1079	(7)	782	21	1861	14
Nov	1081	(5)	760	(1)	1841	(6)
Dec	1075	(12)	763	2	1838	(10)

Table IX display's the monthly seasonal effect in crude oil and petroleum products between the years of 2010 and 2018. The column 'Coef.' display's the dummy coefficients obtained from regressing inventories on twelve seasonal dummy variables. The column 'Seasonality' display's the difference between the corresponding coefficient minus the average coefficient. To clarify with an example, the 'Seasonality' column shows that crude oil inventories tend to fall with 19 million barrels in January, petroleum products tend to fall with 3 million barrels in January and crude oil and petroleum products tend to fall with 22 million barrels in January.

By examining the table some clear seasonal patterns can be found. For instance, during the time period between March and June petroleum products tend to fall the most, while crude oil tend to build the most. This effect is mainly created due to refinery maintenance in the first quarter. The United States rely more on gasoline, as compared to diesel fuel, than most of other countries. U.S. refineries are therefore optimized to produce gasoline, with maintenance schedules placed in accordance to gasoline demand. The demand for gasoline is generally the lowest in January and February, meaning that refinery maintenance is often scheduled during the first quarter of the year. This time also fits in-between the peak heating oil season and the peak summer driving season, allowing refineries to prepare for summer-blend fuels.

Examining crude oil and petroleum products together also demonstrate an interesting seasonal pattern. That is, crude oil and petroleum products tend to fall during the winter months, when the U.S. increase their use of distillate heating oils and residual fuels, while crude oil and petroleum products tend to build during the summer months, as supply generally exceed the demand in the summer.

B. *Price-Inventory Correlation*

Table X display's the correlation between relative inventory and prices, while using different input data for the relative inventory level. As expected the correlation is as strongest when matching U.S inventories with WTI spot prices and OECD inventories with Brent spot prices. Moreover, the price correlation is stronger for crude oil compared to petroleum products. Crude oil and petroleum products are cointegrated processes which mean revert

Table X: Correlation Between Relative Inventory and Prices

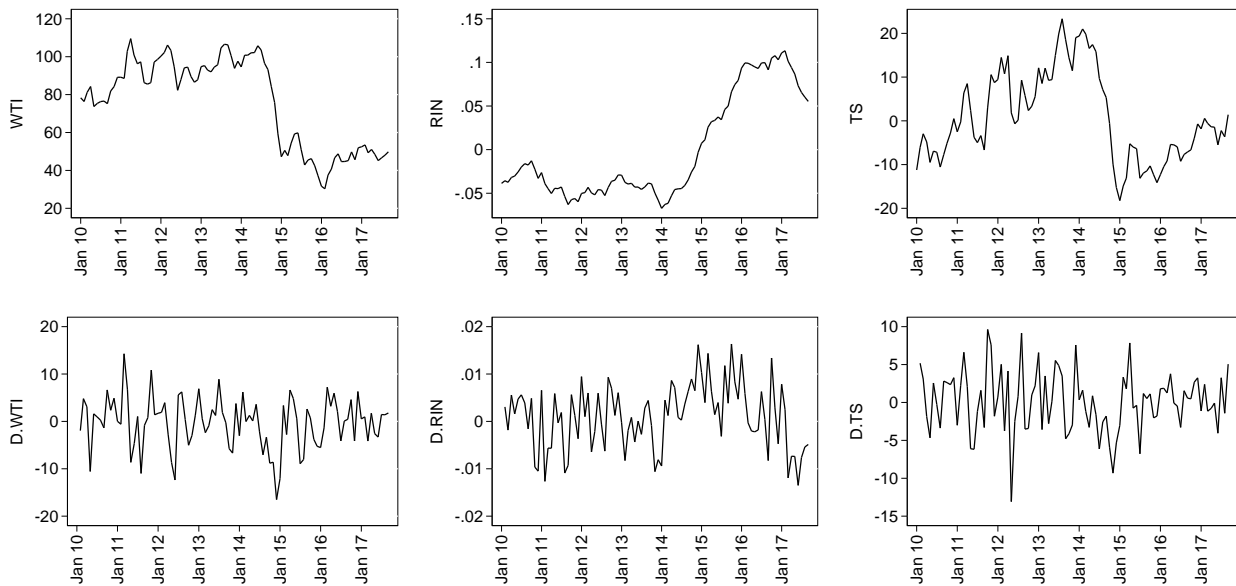
<i>United States</i>	<i>Including SPR</i>		<i>Excluding SPR</i>	
	WTI	Brent	WTI	Brent
Crude & Products	(0.9047)	(0.9032)	(0.9002)	(0.8911)
Crude	(0.8960)	(0.8861)	(0.8864)	(0.8632)
Products	(0.8860)	(0.8939)	(0.8865)	(0.8944)
<i>OECD exc U.S.</i>	<i>Including SPR</i>		<i>Excluding SPR</i>	
	WTI	Brent	WTI	Brent
Crude & Products	(0.8735)	(0.8897)	NA	NA

to each other over time. Petroleum products have higher volatility and deviate more from the mean, contributing to a lower price correlation.

C. ARIMA Identification

The first step of an ARIMA is to determine the order of integration of the variables, which is easily done by visualizing the data and employing the Augmented Dickey-Fuller test.

Figure 7: Stationarity Identification



The Augmented Dickey-Fuller test is used to test for non-stationarity. A rejection of the H_0 , i.e. that there is a unit root at some level of confidence, implies stationarity. All the variables are tested from lag 0 to lag 5. As all the variables are found to be non-stationary in levels and integrated by order one, the ARIMA is differenced once.

The next step is to choose the p and q parameters for the ARIMA model. Identification of the orders of p and q is carried out by comparing the estimated partial and simple autocorrelation of the stationary time series.

Figure 8 demonstrates the partial auto correlation function (PACF) and the simple autocorrelation function (ACF) of the differenced WTI spot price. The horizontal axis represent the differenced lag of the WTI price, the vertical axis represent the autocorrelation of the differenced WTI price and the gray area represent the confidence band. A general tips when identifying the orders of p and q of the ARIMA is to follow the principle of parsimony by differencing at most on time. [Nau \(2005\)](#) gives the following rule of thumb; "in most cases either p is zero or q is zero, and $p + q$ is less than or equal to 3". As the partial auto correlation function displays a positive lag two

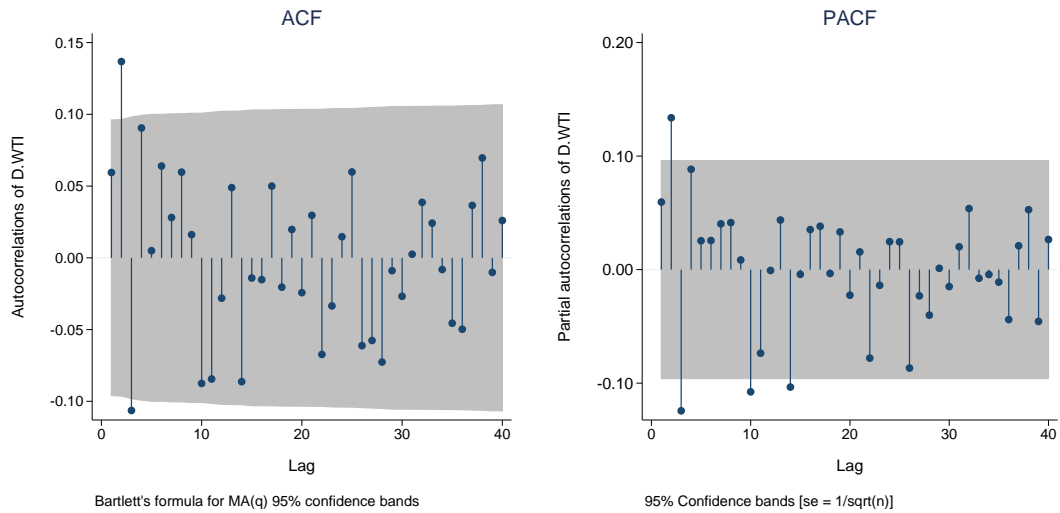
Table XI: Unit-Root Test

Variable	Reject H_0	Test specification	Test Result
WTI	No	No Constant	$I(1)$
WTI	No	Constant	$I(1)$
WTI	No	Constant + Trend	$I(1)$
RIN	No	No constant	$I(1)$
RIN	No	Constant	$I(1)$
RIN	No	Constant + Trend	$I(1)$
TS	No	No Constant	$I(1)$
TS	No	Constant	$I(1)$
TS	No	Constant + Trend	$I(1)$

Table XII: Unit-Root Test

Variable	Reject H_0	Test specification	Test Result
D.WTI	Yes	No Constant	$I(0)$
D.WTI	Yes	Constant	$I(0)$
D.WTI	Yes	Constant + Trend	$I(0)$
D.RIN	Yes	No Constant	$I(0)$
D.RIN	Yes	Constant	$I(0)$
D.RIN	Yes	Constant + Trend	$I(0)$
D.TS	Yes	No Constant	$I(0)$
D.TS	Yes	Constant	$I(0)$
D.TS	Yes	Constant + Trend	$I(0)$

Figure 8: Stationarity Identification



autocorrelation that is followed by a sharp cutoff, the serie appears slightly under differenced. This suggest that I should add one AR term to the model, resulting in an ARIMA (1,1,0).

D. Cross-Validation

Cross-validation is a method to assess predictive models by dividing the original sample into a training set to train the model, and a test set to assess the model. The most basic form of cross-validation is the holdout method, which involves a single run of the data. The data is randomly assigned into two sets, one called the training set and one called the test set. I then proceed to train the model on the training set and evaluate the model on the test set. Although being very intuitive, the holdout method is prone to sample biases. That is, if you change the splitting pattern, there is a chance that the results will be different. More sophisticated versions of cross-validation are the K-fold and the leave-one-out cross-validation (LOOVC) method, which uses multiple runs that are aggregated together. In K-fold cross validation, the original sample is randomly divided into k approximately equal size folds. Of the k folds, a single fold is set aside with the purpose to test the model, while the remaining k-1 folds are used as training data for the model. Lastly, the k-fold cross-validation process is repeated k times, resulting in each separate fold being used exactly once as the validation data.

In LOOVC each learning set is created by taking all the samples except one test sample. I then evaluate the models error on the single-point held out, and repeat this procedure for each of the data points. For n samples, I

therefore have n different training sets and n different test sets (Gilliland et al., 2016). K-fold and LOOVC are not always appropriate for time series as they do not operate within the constraints of the definition of time series models. In K-fold and LOOVC, you split the data into folds and shuffle past values with futures values. Each fold would lose order of sequence and thus lose significance as values occurring over time. Simply, K-fold and LOOVC are appropriate for models with independent and identically distributed data. K-fold and LOOVC are therefore valid for autoregressive models like the ARIMA(p,d,q) which assumes independent and identically distributed data (Bergmeir et al., 2018).

E. Diebold–Mariano Tests

Define actual values as $\{WTI_t; t = 1, \dots, T\}$ and the two competing forecasts as $\{\widehat{WTI}_{1t}; t = 1, \dots, T\}$
 $\{\widehat{WTI}_{2t}; t = 1, \dots, T\}$

Define the forecast errors as: $e_{it} = \widehat{y}_{it} - y_t, \quad i = 1, 2$

The loss attributed to the forecast is assumed to be a function of the forecast error, e_{it} , denoted by $g(e_{it})$. The loss differential between the two forecasts can be stated as $d_t = g(e_{1t}) - g(e_{2t})$. The two predictions will have equal accuracy if and only if the expected value of the loss differential is zero for all time periods.

The test than tests the null hypothesis

$$H_0 : E(d_t) = 0$$

versus the alternative hypothesis

$$H_0 : E(d_t) \neq 0$$

The null hypothesis states that the two forecasts have equal accuracy while the alternative hypothesis states that the two forecasts have different accuracy.

$$\frac{\bar{d} - u}{\sqrt{\frac{2\pi f_d(0)}{T}}} \rightarrow N(1, 0) \quad (16)$$

REFERENCES

- Adhikari, R. and Agrawal, R. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- Allison, P. (2014). Prediction vs. causation in regression analysis. *Statistical Horizons*.
- Alquist, R. and Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25(4):539–573.
- Alquist, R., Kilian, L., and Vigfusson, R. J. (2013). Forecasting the price of oil. In *Handbook of economic forecasting*, volume 2, pages 427–507. Elsevier.
- Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2014). Causality and endogeneity: Problems and solutions. *The Oxford handbook of leadership and organizations*, 1:93–117.
- Bahgat, G. (2016). Lower for longer: Saudi arabia adjusts to the new oil era. *Middle East Policy*, 23(3):39–48.
- Barron, M. (2014). Loocv: Stata module to perform leave-one-out cross-validation.
- Baum, C. (2011). Dmario: Stata module to calculate diebold-mariano comparison of forecast accuracy.
- Baum, C. (2017). Fcstats: Stata module to compute time series forecast accuracy statistics.
- Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.
- Berman, E., Shapiro, J. N., and Felter, J. H. (2011). Can hearts and minds be bought? the economics of counterinsurgency in iraq. *Journal of Political Economy*, 119(4):766–819.
- Bodell, J. M. (2009). The dynamics of natural gas price formation: Implications for gas producers.
- Boons, M. and Prado, M. P. (2015). Basis-momentum in the futures curve and volatility risk.
- Bopp, A. E. and Lady, G. M. (1991). A comparison of petroleum futures versus spot prices as predictors of prices in the future. *Energy Economics*, 13(4):274–282.
- Bouchouev, I. (2012). Inconvenience yield, or the theory of normal contango. *Quantitative Finance*, 12(12):1773–1777.
- Brennan, M. J. (1958). The supply of storage. *The American Economic Review*, 48(1):50–72.
- Byun, S. J. et al. (2017). Speculation in commodity futures markets, inventories and the price of crude oil. *The Energy Journal*, 38(5).
- Chernenko, S., Schwarz, K., and Wright, J. H. (2004). The information content of forward and futures prices: Market expectations and the price of risk.
- Chinn, M. D. and Coibion, O. (2014). The predictive content of commodity futures. *Journal of Futures Markets*, 34(7):607–636.
- Coppola, A. (2008). Forecasting oil price movements: Exploiting the information in the futures market. *Journal of Futures Markets*, 28(1):34–56.
- Daniel, S., Schroeder, T., and Dhuyvetter, K. (1999). Forecasting performance of storable and non-storable commodities. In *NCR-134 Conference: Applied Commodity Price Analysis, Forecasting, and Market Risk Management: Chicago, Illinois, April 19-20, 1999*, page 317.
- Daniels, B. (2012). Crossfold: Stata module to perform k-fold cross-validation.
- de Groot, W., Karstanje, D., and Zhou, W. (2014). Exploiting commodity momentum along the futures curves. *Journal of Banking & Finance*, 48:79–93.
- Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica: Journal of the Econometric Society*, pages 410–421.
- Etienne, X. L. and Mattos, F. (2016). The information content in the term-structure of commodity prices.
- Fama, E. F. and French, K. R. (2016). Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. In *The World Scientific Handbook Of Futures Markets*, pages 79–102. World Scientific.
- Geman, H. and Kharoubi, C. (2008). Wti crude oil futures in portfolio diversification: The time-to-maturity effect. *Journal of Banking & Finance*, 32(12):2553–2559.
- Gilliland, M., Sglavo, U., and Tashman, L. (2016). *Business Forecasting: Practical Problems and Solutions*. John Wiley & Sons.
- Granger, C. W. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2):111–120.
- Hicks, J. R. (1939). *Value and capital*. Oxford At The Clarendon Press; London.
- Jin, X. (2017). Do futures prices help forecast the spot price? *Journal of Futures Markets*, 37(12):1205–1225.
- Kaldor, N. (1939). Speculation and economic stability. *The Review of Economic Studies*, 7(1):1–27.
- Kaufmann, R. K., Dees, S., Karadeloglou, P., and Sanchez, M. (2004). Does opec matter? an econometric analysis of oil prices. *The Energy Journal*, pages 67–90.
- Keynes, J. M. (1930). *A treatise on money: in 2 volumes*. Macmillan & Company.
- Kleinberg, R. L., Paltsev, S., Ebinger, C. K., Hobbs, D., and Boersma, T. (2016). Tight oil development economics: benchmarks, breakeven points, and inelasticities. *MIT Center for Energy and Environmental Research*.

- McCallum, A. and Wu, T. (2005). Do oil futures prices help predict future oil prices?
- Merino, A. and Ortiz, Á. (2005). Explaining the so-called “price premium” in oil markets. *OPEC Energy Review*, 29(2):133–152.
- Moosa, I. A. and Al-Loughani, N. E. (1994). Unbiasedness and time varying risk premia in the crude oil futures market. *Energy economics*, 16(2):99–105.
- Nau, R. F. (2005). Introduction to arima: nonseasonal models. *Duke University*.
- Reeve, T. A. and Vigfusson, R. (2011). Evaluating the forecasting performance of commodity futures prices.
- Reichsfeld, D. A., Roache, S. K., et al. (2011). *Do commodity futures help forecast spot prices?* International monetary fund (IMF).
- Sachs, G. (2015). The new oil order: Lower for even longer.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Shmueli, G. (2010). To explain or to predict? *Statistical science*, pages 289–310.
- Sober, E. (2002). Instrumentalism, parsimony, and the akaike framework. *Philosophy of Science*, 69(S3):S112–S123.
- Telser, L. G. (1958). Futures trading and the storage of cotton and wheat. *Journal of Political Economy*, 66(3):233–255.
- Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of econometrics*, 66(1-2):225–250.
- Working, H. (1949). The theory of price of storage. *The American Economic Review*, 39(6):1254–1262.
- Ye, M., Zyren, J., and Shore, J. (2002). Forecasting crude oil spot price using oecd petroleum inventory levels. *International Advances in Economic Research*, 8(4):324–333.
- Ye, M., Zyren, J., and Shore, J. (2003). Elasticity of demand for relative petroleum inventory in the short run. *Atlantic Economic Journal*, 31(1):87–102.
- Ye, M., Zyren, J., and Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, 21(3):491–501.
- Ye, M., Zyren, J., and Shore, J. (2006). Forecasting short-run crude oil price using high-and low-inventory variables. *Energy Policy*, 34(17):2736–2743.
- Zamani, M. (2004). An econometrics forecasting model of short term oil spot price. In *6th IAEE European conference*, page 2. Citeseer.