Forecasting Monthly WTI Crude Oil Prices

CAUSAL INFERENCE 2020

Mads Emil Marker Jungersen
Study ID: 201906249

Jeppe Dalgaard Rode
Study ID: 201807521

January 4, 2021

1 Introduction

Crude oil is fundamental in residential, commercial and industrial contexts. Thus, movements in crude oil prices have major impacts on the world around us. In 2020, we saw how the price of US oil turned negative for the first time in history. The price of West Texas Intermediate (WTI), the benchmark for US oil, fell as low as minus $ 37.63 per barrel. This led the Organization of the Petroleum Exporting Countries (OPEC) and its allies to cut global oil output by 10% in a record deal [1]. These market conditions have prompted analysts to deem the situation a "forecasting nightmare", especially referencing the many uncertainties evolving the oil market at present. [2].

With oil prices being debated more than ever, this paper aims to investigate the forecasting capabilities of oil prices, especially focusing on how other variables are capable (or incapable) of increasing forecasting performance. As the movements of oil prices are paramount to the global economy, a great amount of research has been dedicated to forecasting oil prices. The methods by which they do so, however vary a great deal. Pindyck (1999)[1] adds a mean reversion term to deterministic trend model, with an implicit ARMA process. This method suggests that oil prices eventually will backslide to the long-run mean, which can apply to not only the asset itself, but also economic growth, for instance. Another approach commonly used to forecast oil prices is ARIMA models, which is a generalization of the aforementioned ARMA models. In this context, one uses difference between values and their preceding values in order to work with non-stationary data. All of these methods have in common that they describe time series data in terms of its own lags. In addition, they are generally not accurate forecasters with regard to oil prices (Hamilton, 2010; Alquist & Killian, 2010)[2, 3].

Hamilton (2010)[2] points to physical market factors, financial market factors and trading factors when explaining why oil forecasts tend to be inaccurate. The most

[1] https://www.bbc.com/news/business-52350082
[2] https://www.cnbc.com/2019/04/11/oil-prices-market-conditions-make-it-almost-impossible-to-forecast.html

successful forecasts have been those who have investigated explanatory factors and then incorporated them into their models. These models have widely found that models based on economic fundamentals (supply and demand, economic growth rate, etc.) work better short-term, whereas models who implement factors with which they are able to estimate the role of speculation and the future behavior of oil supply and demand are more efficient long-term (Miao et al., 2017)[4].

2 Data

The data in this paper consists of several different variabels, collected from different datasets joined together on the date. We use monthly data, because it was the lowest common update frequancy for each of the collected variables.

The dependent variable `Price` is the monthly West Texas Intermediate (WTI) crude oil spot price, and is collected from the U.S. Energy Information Administration (EIA) [3]. The independent or explanatory variables used to help forecast the WTI crude oil price are selcted following EIA's analysis of main factors that drive the oil price.[4] As a meassure for the global supply of crude oil, `Supply`, we use the global production of crude oil including lease condensate meassured in thousand barrels pr. day averaged on a monthly frequancy.[5] To get an estimate of the global demand for oil, which is widely accepted to be associated with fluctuations in the global real economic activity, we use the kilian index `KI`.[6] Kilian and Zhou (2018) [5] argues the power of the kilian index to model global real economic activity. To model the fluctuations in the financial market we use the monthly adjusted closing price of the SP 500 index.[7]

Having collected all the different variables, we join them together on the `Date` variable contained in all of the datasets. The resulting dataset spans from January 1986 to August 2020. In addition, we create multiple geopolitical dummy variables for previous periods, that has been known to have an influence on the WTI crude oil price.[8] We define the following variables:

- `IIK` for when Irak invaded Kuwait, and takes the value 1 in the period from August to September in 1990.
- `AFC` for the asian financial crisis and takes the value 1 in the period from July 1997 to December 1998.
- `NE` for the 9/11 attack, and takes the value 1 in the period from September to November 2001.
- `GFC` for the global financial crises, and takes the value 1 in the period from September 2008 to February 2009.

[3] https://www.eia.gov/dnav/pet/hist/rwtcM.htm
[4] https://www.eia.gov/finance/markets/crudeoil/
[5] https://www.eia.gov/opendata/qb.php?sdid=INTL.57-1-WORL-TBPD.M
[6] https://www.dallasfed.org/research/igrea
[7] Data found at https://finance.yahoo.com
[8] https://www.eia.gov/finance/markets/crudeoil/spot_prices.php

Additionally we added monthly dummy variables, to be able to test for seasonality in the data.

3 Methodology

In this paper, we employ the method of autoregressive models (AR) - using OLS - in order to forecast WTI Crude oil prices. As OLS only is asymptotically valid when the dependent and independent variables are stationary and weakly dependent, we implement the augmented Dickey-Fuller test (ADF) which checks whether a time series has a unit root. The ADF enables us to determine if a time series is stationary or non-stationary, with the latter indicating that first differencing is necessary in order to obtain stationarity.

Autoregressive models, in general, make use of preceding values and uses these to predict future values. We choose to work with real prices, and not the logarithmic differences, so that we elude the log approximation error when forecasting the prices of oil (p. 663; Wooldridge). We will work with different autoregressive models of different orders. Further, as we aim to investigate what factors that influence the price of WTI Crude oil, we also choose to incorporate other externals factors that may affect the forecast, changing our model such that it will consider these particular factors. This model is known as AR(1)+X, where X stands for other exogenous variables and implies that the output variable y_t depends linearly on its own preceding values, a stochastic term and whichever lagged exogenous variables we may choose to incorporate. We use test for granger causality to determine the number of lags for the dependent variable, as well as to choose the explanatory variables to use in our model.
In order to evaluate the quality of our forecasts, we use out-of-sample performance meassures. This approach entails that we first subset our data into two groups: a training set and a test set. Then, we give each model the training-data in order to estimate its parameters and use our training-data for out-of-sample testing. Lastly, we evaluate our models' forecasting capabilities and performances, comparing them to a random walk as a benchmark.

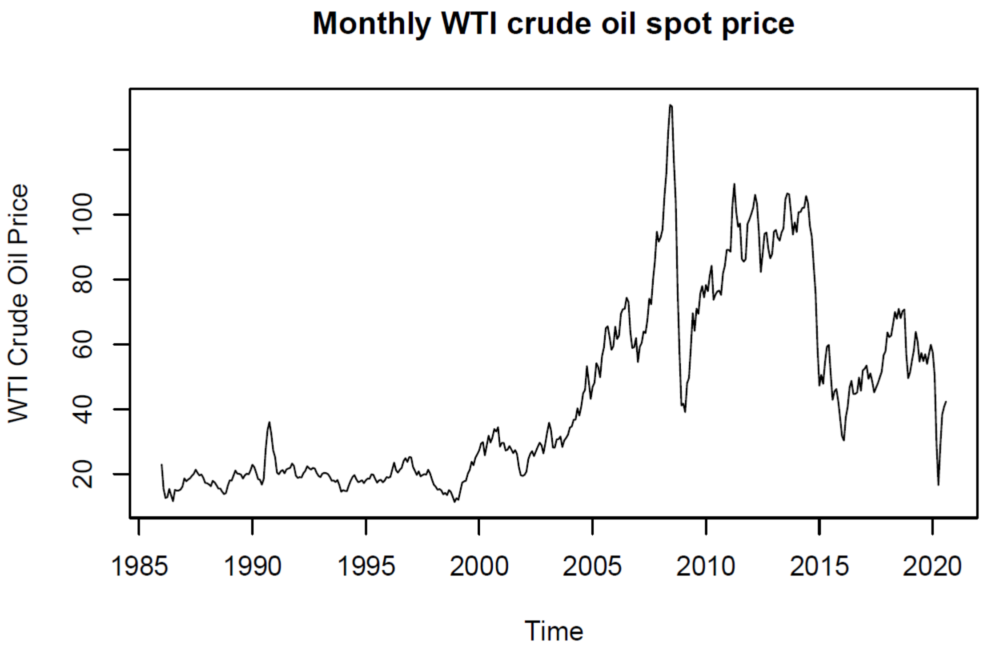
We subset our data such that our training-data spans from January 1986 to December 2013 and our test-data spans from January 2014 to August 2020, which roughly corresponds to an 80/20 % split of our data. The root mean squared (RMSE) and the mean absolute error (MAE) of the predictions are the measures with which we estimate how accurate our predictive model is with regard to our test-data, implying that the lower the values of the out-of-sample RMSE and MAE the stronger the predictive model (p. 659; Wooldridge). The particular aspect of our forecast performance that we aim to evaluate is the one-step-ahead performance. In this approach, we evaluate the following error:

$$\hat{e}_{n+1} = y_{n+1} - \hat{f}_n$$

Here, we use the estimated parameters along with the prices that precede the one being predicted. We end our analysis by testing for serial correlation in our forecasting models. We use the Breusch-Godfrey (BG) test, testing for $AR(1)$ serial correlation where we do not assume strict exogeneity. We compare the p-value with a Durbin Watson (DW) test, that is valid for finite samples, but do however assume strictly exogenous regressors. We also test wheter our models suffer from heteroskedasticity using a Breuch Pegan (BP) test, that assumes no serial correlation.

4 Empirical results

As a fist step in our analysis, we analyze the stationary properties of the variables of interest. Looking at the plot of the monthly WTI spot prices from January 1986 to August 2020, we observe no indication of a stationary process.



We can make similar plots for the other variables, to visually explore the stationarity properties. More formally using the ADF test to test for stationarity yields the following results:

Augmented Dickey Fuller test			
	# Lags	Test Statistic	P-value
Price	7	-2.129	0.523
Supply	7	-0.887	0.954
KI	7	-2.628	0.312
SP500	7	0.027	0.990

The number of lags included in the test is automatically set by the `adf.test()` function in R. We observe that the ADF produces p-values over 0.05 for all variables, and hence we cannot reject the null hypothesis of unit root for those variables, implying that these variables are likely $I(1)$ processes. In order to account for that we use first differencing to make those variables weakly dependent and stationary, which can be verified with an ADF test on the differenced variables (p. 396; Wooldridge).

With stationary variables we now define and estimate the forecasting models. Estimating the $AR(1)$ model on the training data-set using the first differenced price variable $\Delta\texttt{Price}$ variable gives:

$$\widehat{\Delta\text{Price}_t} = \underset{(0.209)}{0.166} + \underset{(0.051)}{0.376}\Delta\text{Price}_{t-1} \tag{1}$$
$$n = 334, \overline{R}^2 = 0.1402$$

Where the intercept is not statistically significant, but the coefficient on the lagged variable is. Adding more lags of $\Delta\texttt{Price}$ to the model show that only $\Delta\texttt{Price}_{t-1}$ and $\Delta\texttt{Price}_{t-6}$ are statistically significant (using regular t-tests), and the lagged terms from $t-2$ to $t-6$ are jointly significant (using regular F-test). We therefore include all 6 lags.
When testing for which af the lagged independent variables that help predict $\Delta\texttt{Price}_t$ after we have controlled for past values of $\Delta\texttt{Price}$, only $\Delta\texttt{SP500}_{t-1}$ do. After controlling for past values of $\Delta\texttt{Price}$ conditional on $\Delta\texttt{SP500}_{t-1}$, none of the remaining variables granger causes $\Delta\texttt{Price}$. We end in the following equation:

$$\begin{aligned}\widehat{\Delta\text{Price}_t} &= \underset{(0.207)}{0.165} + \underset{(0.055)}{0.314}\Delta\text{Price}_{t-1} + \underset{(0.058)}{0.061}\Delta\text{Price}_{t-2} - \underset{(0.058)}{0.069}\Delta\text{Price}_{t-3} \\ &\quad - \underset{(0.058)}{0.018}\Delta\text{Price}_{t-4} - \underset{(0.058)}{0.045}\Delta\text{Price}_{t-5} - \underset{(0.055)}{0.157}\Delta\text{Price}_{t-6} + \underset{(0.005)}{0.016}\Delta\text{SP500}_{t-1} \\ n &= 329, \overline{R}^2 = 0.199, \text{ Model-Name } = \text{AR6X}\end{aligned} \tag{2}$$

For comparison reasons we also define a model only including 1 lag of the differenced oil price. Using similar tests, we agian find that only $\Delta\texttt{SP500}_{t-1}$ help predict $\Delta\texttt{Price}_t$ when we have controlled for past value of $\Delta\texttt{Price}$.

$$\widehat{\Delta\text{Price}_t} = \underset{(0.207)}{0.086} + \underset{(0.05)}{0.348}\Delta\text{Price}_{t-1} + \underset{(0.005)}{0.018}\Delta\text{SP500}_{t-1} \tag{3}$$
$$n = 334, \overline{R}^2 = 0.171, \text{ Model-Name } = \text{AR1X}$$

Running exactly the same tests as in the two models above, but now controlling for previous events that have influenced the crude oil price reaches the same conclusion (only $\Delta\texttt{SP500}_{t-1}$ helps predicting $\Delta\texttt{Price}_t$), however the coefficients on the variables changes.

$$\begin{aligned}\widehat{\Delta\text{Price}_t} &= \underset{(0.209)}{0.566} + \underset{(0.056)}{0.158}\Delta\text{Price}_{t-1} - \underset{(0.057)}{0.053}\Delta\text{Price}_{t-2} - \underset{(0.055)}{0.137}\Delta\text{Price}_{t-3} \\ &\quad - \underset{(0.055)}{0.056}\Delta\text{Price}_{t-4} - \underset{(0.054)}{0.041}\Delta\text{Price}_{t-5} - \underset{(0.052)}{0.143}\Delta\text{Price}_{t-6} + \underset{(0.005)}{0.012}\Delta\text{SP500}_{t-1} \\ &\quad + \underset{(2.483)}{5.910}\text{IIK}_t - \underset{(0.85)}{1.347}\text{AFC}_t - \underset{(2.037)}{2.469}\text{NE}_t - \underset{(1.931)}{12.430}\text{NE}_t \\ n &= 329, \overline{R}^2 = 0.30, \text{ Model-Name } = \text{AR6Xd}\end{aligned} \tag{4}$$

And

$$\widehat{\Delta\text{Price}_t} = \underset{(0.209)}{0.342} + \underset{(0.056)}{0.186}\Delta\text{Price}_{t-1} + \underset{(0.005)}{0.013}\Delta\text{SP500}_{t-1} + \underset{(2.558)}{6.439}\text{IIK}_t - \underset{(0.872)}{0.935}\text{AFC}_t$$
$$- \underset{(2.098)}{1.982}\text{NE}_t - \underset{(1.755)}{9.454}\text{GFC}_t \tag{5}$$
$$n = 334, \overline{R}^2 = 0.244, \text{ Model-Name } = \text{AR1Xd}$$

Comparing the in-sample forecasting performance meassured in the adjusted $\overline{R}^2$ value, the `AR6Xd` model do a better job, however we are more interested in the out-of-sample forecasting. For all of the above estimated models we now evaluate the forecasting performance using the out-of-sample RMSE and MAE for the remaing 20% of the dataset. For for comparison reasons we include the Random Walk (without a drift) as a benchmark for our results. Furthermore we add the $AR(1)$ and $AR(6)$ model to see if Δ`SP500`$_{t-1}$ actually do help predict the differenced oil price.

Forecassting performance							
	AR6X	AR1X	AR6Xd	AR1Xd	AR(6)	AR(1)	Random Walk
RMSE	5.247	5.146	5.377	5.226	5.527	5.387	5.796
MAE	4.233	4.127	4.141	4.063	4.276	4.146	4.340

We also include the performance of the models when holding out data for 2020 to see if the same models performs best.

Forecassting performance leaving out 2020							
	AR6X	AR1X	AR6Xd	AR1Xd	AR(6)	AR(1)	Random Walk
RMSE	4.917	4.897	5.050	4.854	4.774	4.694	4.958
MAE	4.013	3.966	3.912	3.832	3.869	3.812	3.862

For the specified models in this reasearch we now test whether any of the models suffer from serial correlation or/and heteroskedsticity.

Tests for serial sorrelation and heteroskedasticity (p-vals)						
	AR6X	AR1X	AR6Xd	AR1Xd	AR(6)	AR(1)
BG	0.300	0.536	0.414	0.044	0.444	0.465
DW	0.546	0.579	0.508	0.183	0.492	0.575
BP	0.003	0.005	0.000	0.000	0.000	0.000

We observe that the BG and DW tests produce high p-values for all models except the `AR1Xd`, which indicate that the null hypothesis of no serial correlation cannot be rejected in these models. For the `AR1Xd` model, the p-value for the BG test is very close to the 0.05 significance level. As we observe no serial correlation (on a 0.01 significance level) in the errors, we are able to implement BP test with which we test for heteroskedasticity of errors in our regression. The null hypothesis for the

BP is that the error variances are all equal. If we test for heteroskedasticity with a significance level of 0.05, then we fail to reject the H_0 hypothesis in all of the cases, suggesting a strong evidence of heteroskedasticity in our models. Note that values less than 0.0005 are just reported as 0.000.

5 Discussion

Our analysis suggest that $\Delta\texttt{SP500}_{t-1}$ help predict future values of $\Delta\texttt{Price}$ when looking at the complete testing period from January 2014 to August 2020. All of the models including the variable $\Delta\texttt{SP500}_{t-1}$ outperform the models without that variable meassured in both RMSE and MAE. When we on the other hand leave out data for 2020 these models are no longer in favour. In this period the simple $AR(1)$ model has the lowest MAE and RMSE, and the AR6Xd model actually predicts worse than the Random Walk benchmark. This difference may be due to the fact that the WTI crude oil price experienced a huge decline in mid-2014 to early 2015. This drop in WTI crude oil price was primarely due to supply factors[9] and hence the SP 500 index didn't experience the same drop. This can be validated by looking at the price development of the two objects in the given time period. This may explain why simple univariate models perform better. In early 2020 both the WTI crude oil price and the SP 500 index experienced a sudden decrease, and therefore the $\Delta\texttt{SP500}_{t-1}$ may be usefull in forecasting the WTI crude oil price.

Looking at the performance of our forecasting models on the entire forecasting period we notice that including dummies for previous geoplitical periods cause the models to have a higher RMSE, but a smaller MAE. The coefficients on the dummy variables is quite high, especially compared the the other coefficients. The reason for this, is that these periods have caused shocks in the WTI crude oil price, and when we do include the dummy variables these shocks are "allocated" to the coefficients on these dummy variables. When we do out-of-sample forecasting these coefficints are set to zero, so we might expect that when we forecast volatile periods, our models including the dummy variables doesn't capture these shocks very well. On the other hand we might expect that under normal not very volatile periods this model predicts well. This might explain why we get reductions in MAE, but higher RMSE because RMSE penalises bigger errors compared to MAE. The forecasting period is a quite volatile period in the WTI crude oil price.

The resulting models was found by carrying out tests for which of the explanatory variables that granger causes the differenced WTI crude oil price, i.e. that help predict future values of $\Delta\texttt{Price}$. We found that only $\Delta\texttt{SP500}_{t-1}$ granger causes $\Delta\texttt{Price}_t$, and the variables Supply and KI was therefore not included in any of our models. This conclusion was reached based on several t- and F-tests, and therefore we need the assumptions that secure the asymptotic properties of OLS. Many of these assumptions seem reasonable in our models, especially the stationary and

[9]https://blogs.worldbank.org/developmenttalk/what-triggered-oil-price-plunge-2014-2016-and-why-it-failed-deliver-economic-impetus-eight-charts

weakly dependent part since our originally unit root variables are first differenced before used in the regression analysis. If we furthore believe that our regressors are contemporaneously exogenous then we may assume that our coefficients are consistent. However more interestingly we found that our models contain no serial correlation at a 0.01 significance level, but all our models suffer from heteroskedasticity. This is important because then our usual standard error formula is no longer valid, and hence the usual t- and F-tests can not be trusted. We should instead have used robust standard-errors, more specifically the Newey-West HAC standard errors, and then carry out our test based on those. This could potentially lead to different conclusions, and could indicate that other variables than only $\Delta\texttt{SP500}_{t-1}$ granger causes $\Delta\texttt{Price}_t$.

It seems counter-intuitive to fundamental economic theory that neither oil supply nor oil demand serve to help forecasting oil prices. However, it is well-documented that oil prices are forward looking. And recent research suggests that there seem to be informational frictions in the oil market between agents. (Killian & Zhou, 2018)[5]. Thus, not only current supply and demand matter, but also news shocks to supply and demand. In this sense, it is important to note the structure of the oil market. The oil market has a peculiar structure, with OPEC dominating the market. OPEC provides frequent updates on its production plans. These updates are closely monitored by markets and can possibly lead to volatile market reactions.

In a forecasting context, the forward-looking properties of oil prices implicate our models. By definition, we use lagged values in our autoregressive models, also with regard to supply and demand. And while these may very well be fundamental drivers of oil prices, expectations may be the proximate cause of their fluctuations. (Sockin & Xiong, 2015)[6]. This indicates that our assumption of contemporaneous or lagged links between oil prices and supply and demand may be flawed. Instead, variables that contain information on the expectations of oil supply and demand seem desirable. Financial factors, such as investors' risk appetites, have been widely used as means to modelling shocks in expectations with regard to supply and demand. In this context, net positions have been used as an indication of expectations (Fueki et al., 2018)[7]. A suggestion for future research may be creating an instrument with which we can directly model the informational frictions in the markets, thus enabling us to identify oil supply and demand news shock more efficiently. With the ongoing technological improvements in machine learning, a neural network that in some way quantifies announcements and informational frictions between oil market participants, for instance on social media, could be an interesting approach to investigate.

In our analysis, we use simple t-tests and F-test to determine which variables were significant with regard to forecasting performance. However, this seems like an arbitrary process, especially considering that we have little to no experiences with economic research. Instead, regression analysis methods that serve to improve forecasting performance, such as LASSO or Boruta, could have been employed for variable selection and regularization.

None of our models contain a trend term nor seasonal dummies. A t-test using the Newey-West HAC standard errors for adding a linear trend term in the models AR6X,AR1X, AR6Xd and AR1Xd all show that the linear trend are not statistically significant on a 0.01 significance level. Also using the Newey-West HAC standard errors in a wald-test for joint significance of the seasonal dummies have p-values way above 0.05, so there is no evidence that seasonality needs to be accounted for in forecasting ΔPrice.

We use one-step-ahead forecasting with model estimation based on a fixed window, namely the period from January 1986 to December 2013. However, it has been well documented that the predictive ability of the forecast parameters on crude oil prices varies over time as analyzed by Kruse and Wegener (2020)[8]. We also saw earlier that the shock in the WTI crude oil price in mid-2014 was due to supply-factors. If that period was within our estimation period we might have chosen to include the Supply variable. This could suggest that we should use a rolling window estimation, conducting tests at each estimation period to see which of our variables granger causes ΔPrice. On the other hand this is computationally more costly, and can lead to overfitting problems. As mentioned the LASSO method is a more sophisticated method of regression that accommodate the overfitting problem by shrinking coefficients without predictive power to zero.

Like written in the introduction models that is based on economic fundamentals work better in the short-term forecasting, at least compared to the Random Walk. Miao et al., (2017)[4] found that a model based on economic fundamentals, where similar but a few more variables were used compared to our analysis, estimated on a 5 year rolling window outperformed, like in our case, the Random Walk in a one month ahead forecasting meassured in the Mean Squared Error. Furthermore they found that the model could be further improved by using the LASSO-method to choose important variables. Instead of assuming a linear relationship between the dependent variable and the regressors, Natarajan and Ashok (2018)[9] creates a relatively simple multilayer perceptron neural network consisting of two hidden layers with 12 neurons each. Using an Exponential Linear Unit activation function, it compares on par with different ARIMA models meassured in both RMSE and MAE.

References

[1] Robert S. Pindyck (1999). The long-run evolutions of energy prices. *The Energy Journal*, 20:1–27.

[2] James D. Hamilton (2010). Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics*, 15.

[3] Ron Alquist and Lutz Kilian (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25:539–573.

[4] Hong Miao, Sanjay Ramchander, Tianyang Wang, and Dongxiao Yang (2017). Influential factors in crude oil price forecasting. *Energy Economics*, 68:77–88.

[5] Lutz Kilian and Xiaoqing Zhou (2018). Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance*, 88:54–78.

[6] Michael Sockin and Wei Xiong (2015). Informational frictions and commodity markets. *The Journal of Finance*, 70:2063–2098.

[7] Takuji Fueki, Hiroka Higashi, Naoto Higashio, Jouchi Nakajima, Shinsuke Ohyama, and Yoichiro Tamanyu (2015). Identifying oil price shocks and their consequences: the role of expectations in the crude oil market. *BIS Working Papers*, 725.

[8] Robinson Kruse and Christoph Wegener (2020). Time-varying persistence in real oil prices and its determinant. *Energy Economics*, 85:104328.

[9] G. Natarajan and Aishwarya Ashok (2018). Multivariate forecasting of crude oil spot prices using neural networks. *ArXiv*, abs/1811.08963.