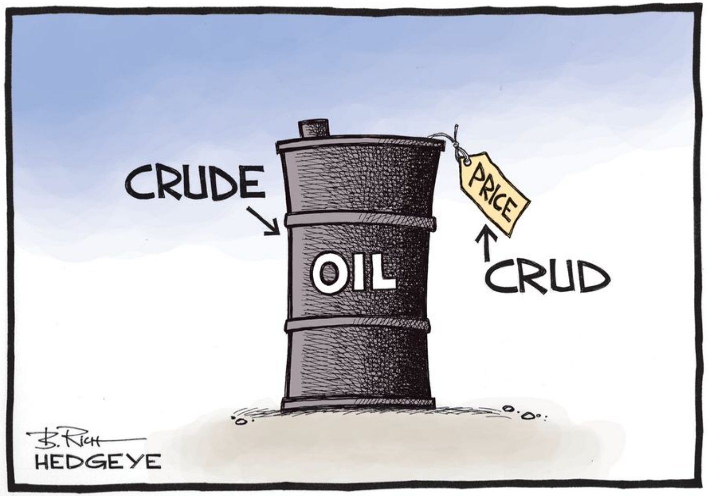
Forecasting Crude Oil Prices



Efforts By: Cade Foster, Mallika Chandra, Quentin Bidwell, Tripti Agarwal

Contents

Introduction and Data Exploration

Introduction

First look at the data:

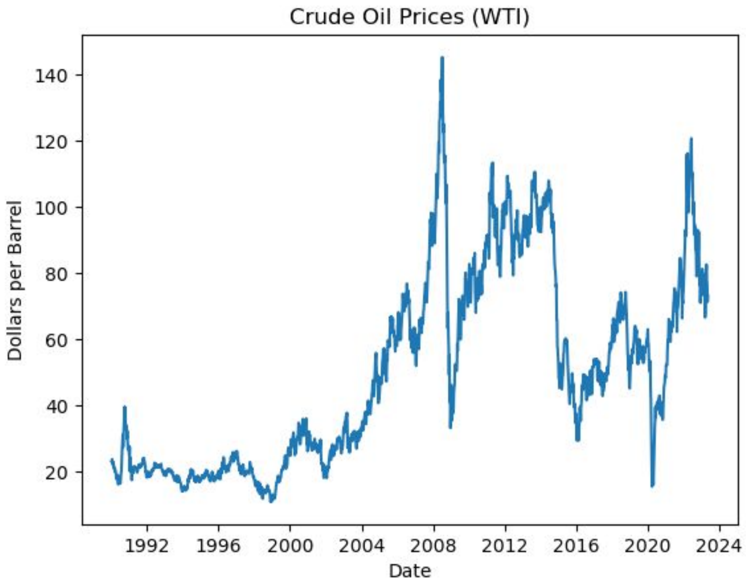
- Crude oil prices, units dollars per barrel, not seasonally adjusted
- Data from St. Louis Fred (1990-2023)
- Cleaned to weekly frequency

Initial thoughts:

- Energy, especially oil, will be highly sensitive to geopolitics
- Forecasts may be undermined by economic shocks

Why it matters?

- Valuable in every industry and yet a challenge to model



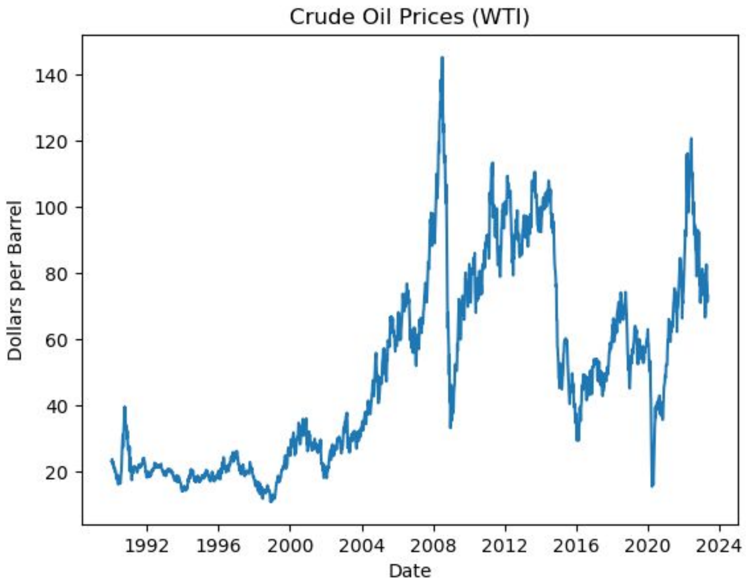
More on data and motivation

Further discussion:

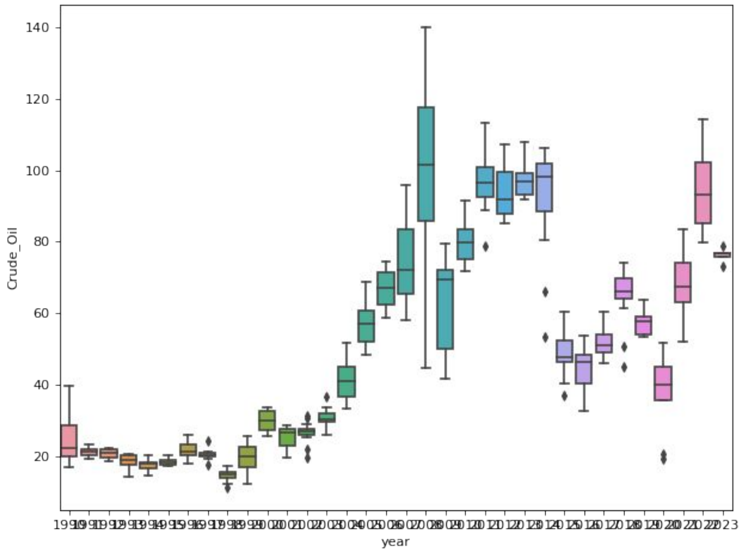
For crude oil prices we refer to the WTI (West Texas Intermediate) prices which serve as a widely recognized benchmark for oil prices in the United States, reflecting the cost of a barrel of oil traded on the New York Mercantile Exchange (NYMEX).

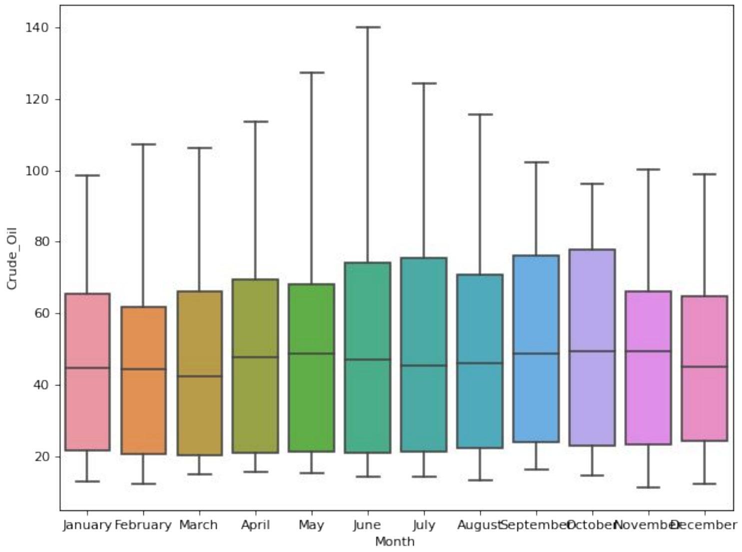
Crude oil prices are influenced by a multitude of factors such as supply and demand dynamics, geopolitical events, and macroeconomic trends. Forecasting crude oil prices can provide valuable insights into the broader energy market dynamics and help anticipate market trends and shifts.

Another aspect that attracted us to this problem was to test the limits of univariate forecasting models on series that intuitively would benefit from a multivariate model because of their sensitivity to economic shocks.



Boxplots of data seasonality





First look at the data:

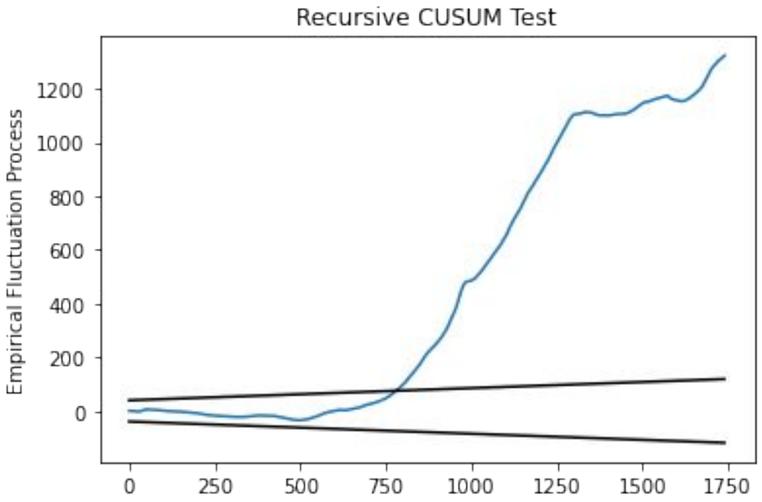
Looking at the boxplots of seasonal factors, at the annual level it's difficult to identify consistent seasonal patterns. For the monthly level, the seasonal pattern is more easily observable albeit somewhat weak.

CUSUM TEST

Interpretation of the structural break:

Clearly the data has a significant departure from the cusum threshold. A spike outside the confidence bands suggests a significant shift or change in the underlying dynamics of the series. This might include a sudden and substantial shift in the mean, variance, or other statistics. It could be caused by various factors such as policy changes, economic events, or external shocks affecting the system being modeled.

This is consistent with our economic expectation of the data being sensitive to economic shocks and reinforces the idea that our analysis will be a test of the limits of univariate forecasting methods on series like these.

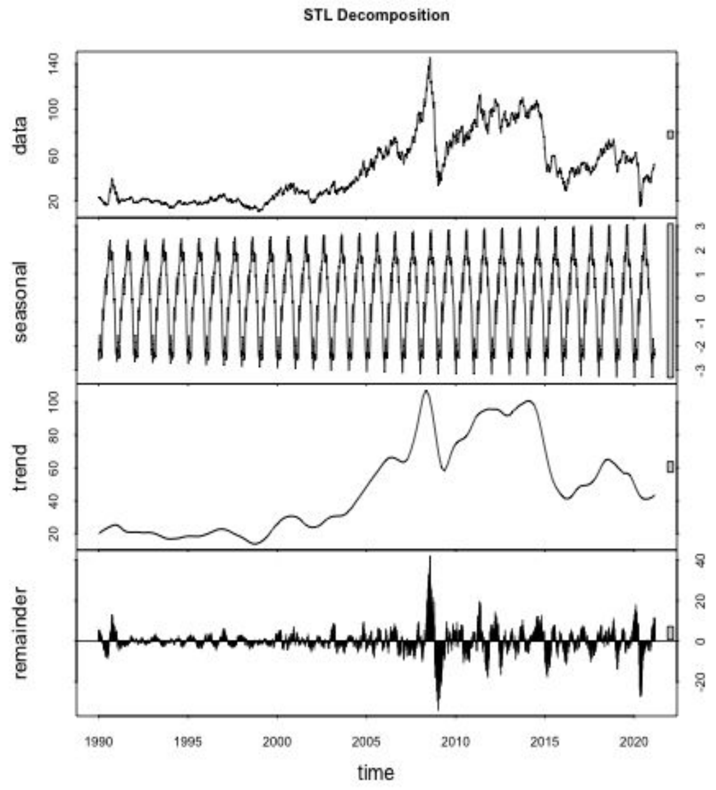


STL DECOMPOSITION

Seasonality: Multiplicative seasonal component as we can see the amplitude of the fluctuations to be mildly increasing. While the multiplicative factor isn't particularly pronounced, it still is the best description of the series's behavior over time.

Trend: No visible upward or downward trend.

Remainder: Centered around zero but with inconsistent dynamics. The visible volatility will be something we explore with GARCH modeling later on.

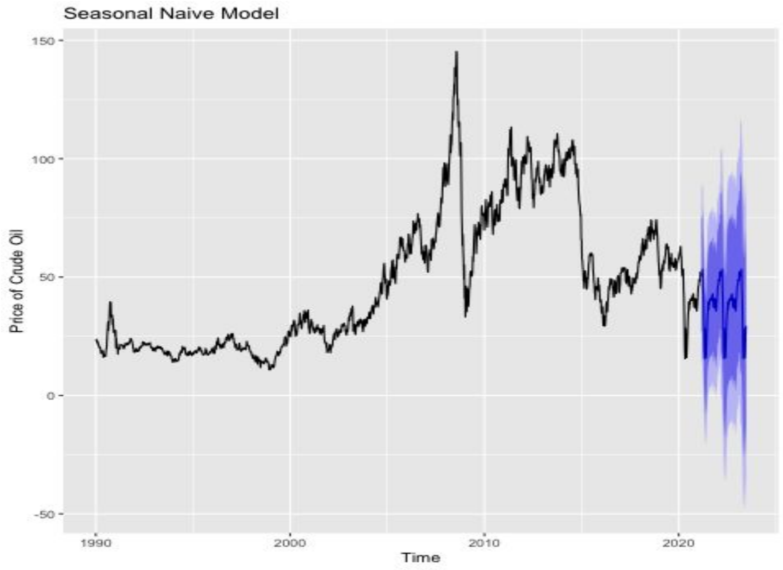


Fitting Different Models

Model 1: Seasonal Naïve

Intuition: The forecasted price of crude oil for a specific week in the future will be the same as the crude oil price from the same week in the previous year

Observation: The model captures only the seasonal component, hence has a high RMSE and MAPE, this is intuitive since we only utilise this as a baseline model



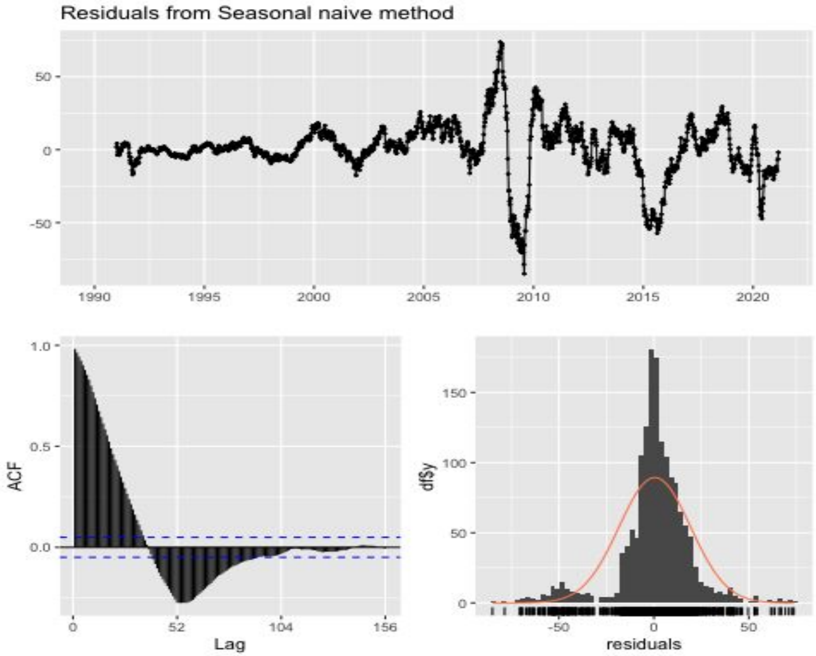
RMSE: 47.791

MAPE: 51.676

Model 1: Seasonal Naïve Residuals

Observation: Based on the residuals we observe that the model does not fare well as the errors are not centred around 0, and the p-value for the Ljung Box Test is lower than the 5% threshold implying that the residuals are not stationary and hence not white noise.

Takeaway: Since the residuals are not white noise, this implies that there are residual components in the model which are not being captures. Hence, the model needs to be improved.

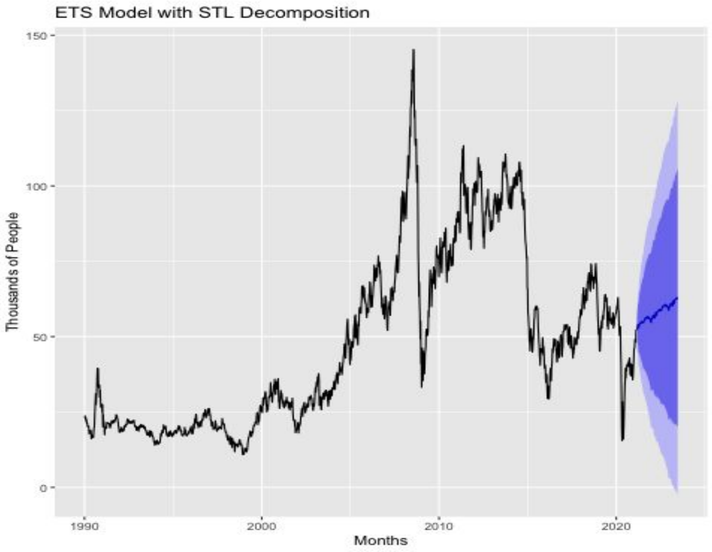


Ljung Box Test
p-value < 2.2e-16

Model II: STL-ETS

Intuition: ETS cannot handle weekly frequency alone, hence we employ STL-ETS to decompose the time series into trend, seasonal, and residual components using the STL algorithm and then apply exponential smoothing to forecast each component separately.

Observation: The forecast has improved and is now capturing the trend and error component as well.



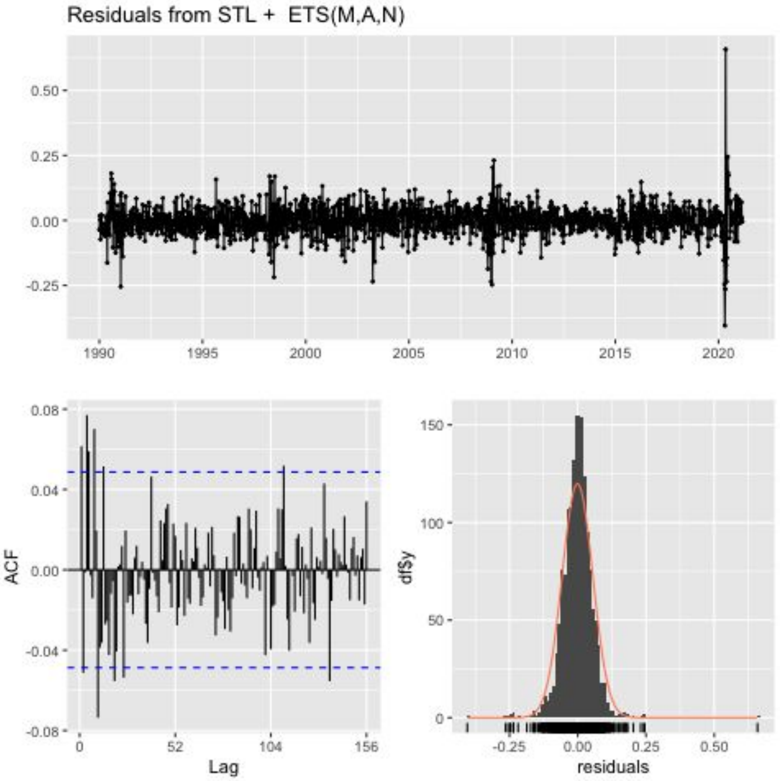
RMSE: 32.039

MAPE: 32.728

Model II: STL-ETS Residuals

Observation: Based on the residuals we observe that the models fares largely well as the errors are centred around 0, and the p-value for the Ljung Box Test is greater than the 5% threshold implying that the residuals are stationary and hence white noise.

Takeaway: This propels us to conclude that the STL-ETS model having a multiplicative error term, additive trend, and no seasonality is able to capture the varying components of the time series successfully.

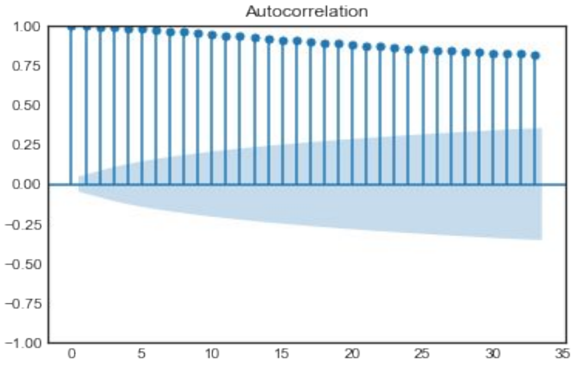


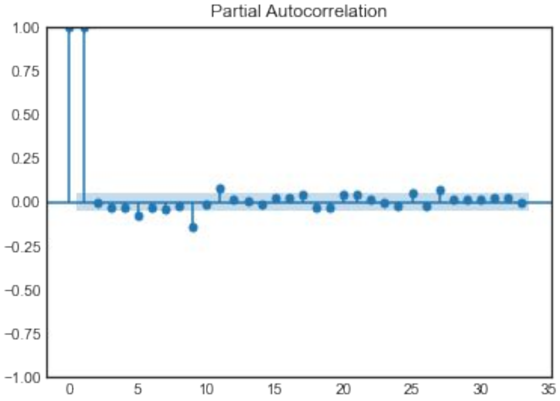
Ljung Box Test
p-value = 0.1475

Model III: ARIMA

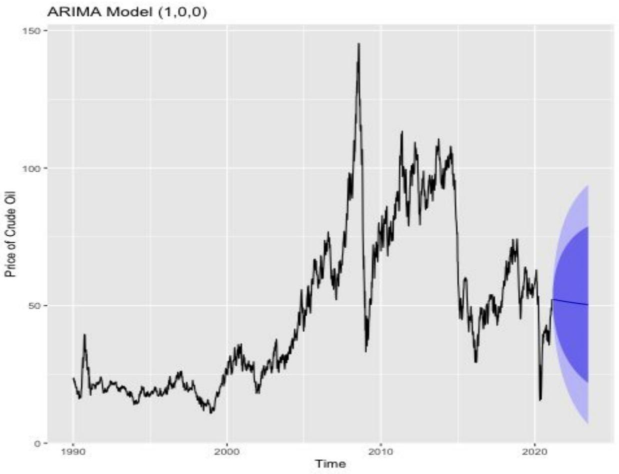
Intuition: It considers the AR component to capture the linear relationship between the current observation and the previous observations, and the MA to account for the influence of past errors or residuals, and the differencing component (I) to remove any trend or seasonality present in the data.

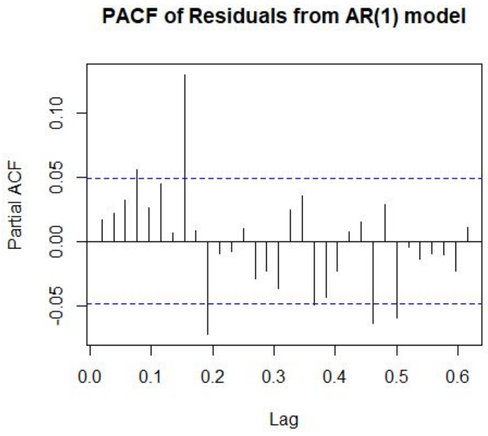
ACF and PACF: From the plots we can conclude that this is an AR(1) process

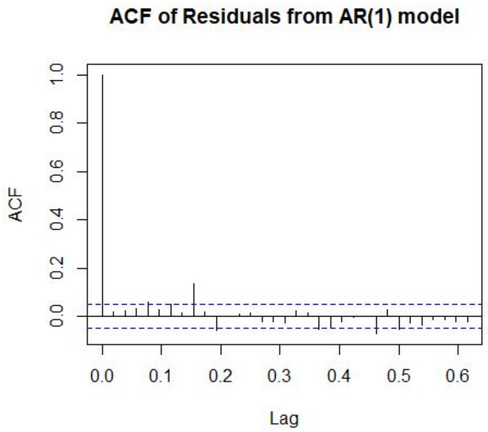




Model: ARIMA





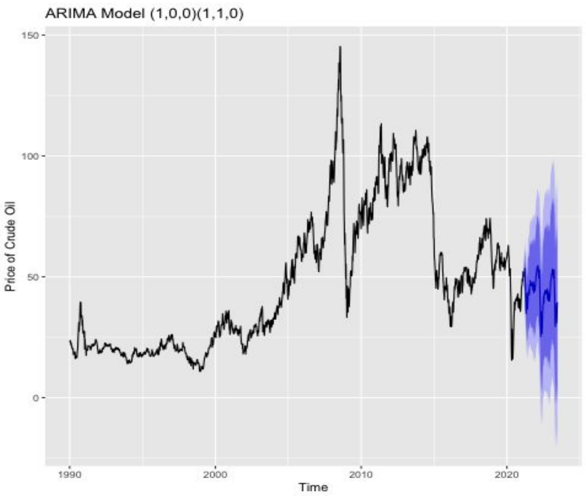


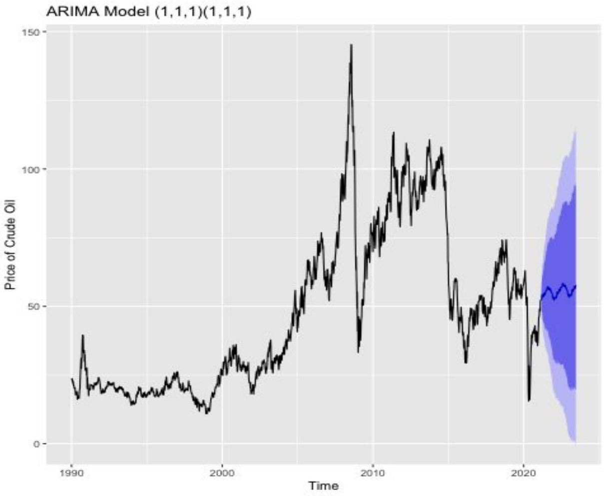
Order: (1,0,0)

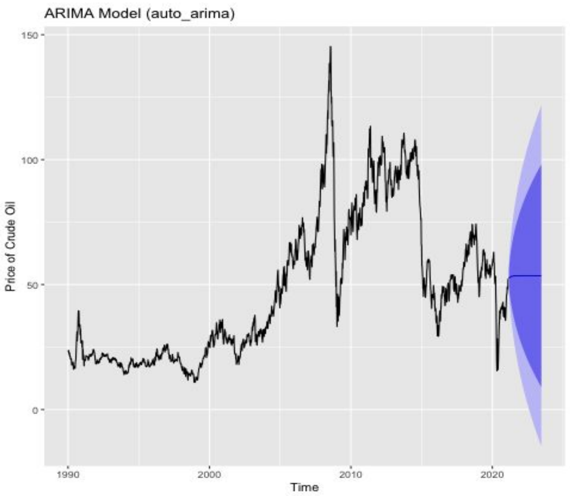
Ljung Box Test: p-value is less than threshold, hence residuals are not stationary

Residual Diagnostics:
Sharp cut-off in ACF, and decay and seasonality in PACF show that some dynamics still remain, motivating a seasonal ARMA model

Model: ARIMA







Order: (1,0,0)(1,1,0)

Ljung Box Test: p-value is less than threshold, hence residuals are not stationary

Order: (1,1,1)(1,1,1)

Ljung Box Test: p-value is less than threshold, hence residuals are not stationary

Order: (1,1,1)

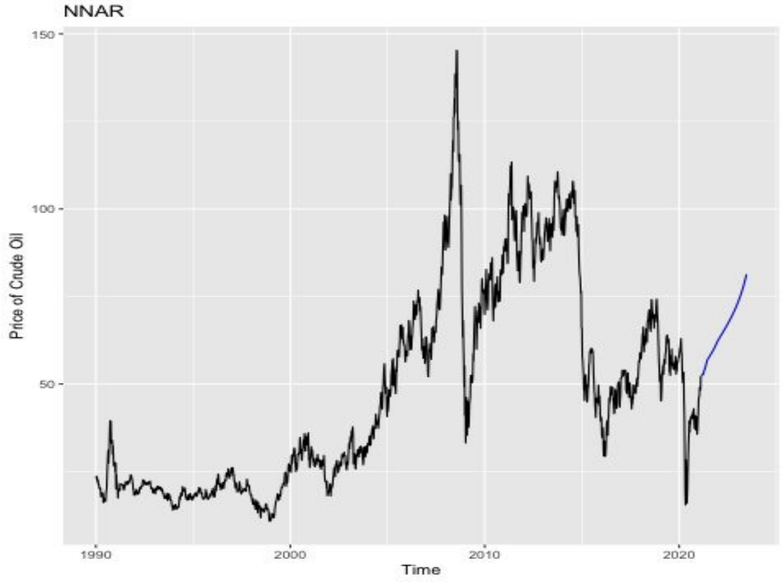
Ljung Box Test: p-value is less than threshold, hence residuals are not stationary

Model IV: NNAR

Intuition: It builds a neural network autoregression model that uses lagged values of the series as inputs to a neural network. It is a more advanced AR model that does away with the stationarity requirement of the data. Since this includes a hidden layer, it helps to model non-linear relationships.

Observation: The optimal model NNAR(11,1,6)[52] implies using 11 lags, 1 seasonal lag and 6 neurons in the hidden layer.

This optimal model was achieved after experimenting with higher seasonal lag orders and multiple repeat parameters. **code in appendix



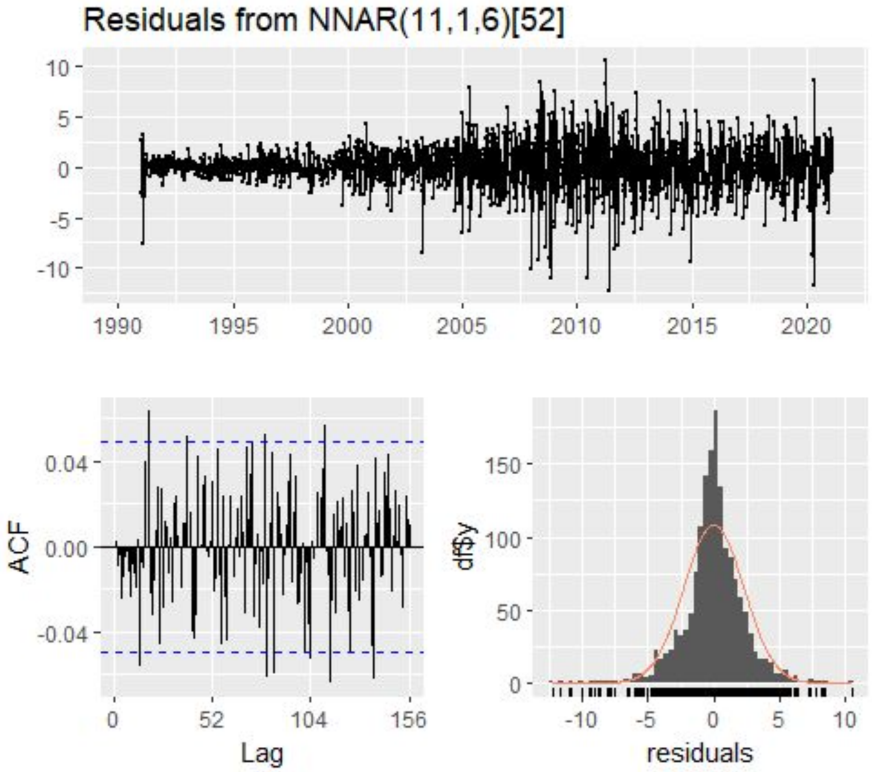
RMSE: 21.647

MAPE: 18.734

NNAR Residuals

Observation:

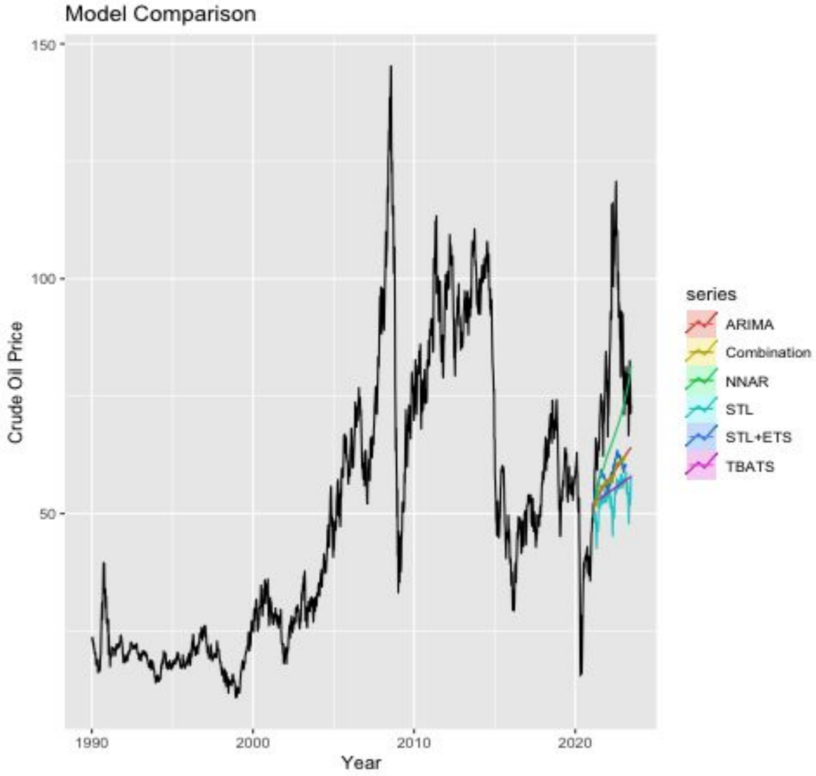
- The optimal NNAR (11,1,6)[52] model reduces the RMSE drastically when compared to ARIMA and auto ARIMA models.
- The residuals closely resemble a white noise process with insignificant AR lags.
- The residuals still exhibit some cyclicity and the first plot on the right shows non-constant volatility, motivating ARCH/GARCH modeling.



Model Comparison

Model	RMSE	MAPE
Naïve	47.791	51.676
STL-ETS	32.039	32.729
ARIMA	27.428	26.376
NNAR	**21.648**	**18.735**
TBATS	30.526	30.402
Combined	29.021	24.336

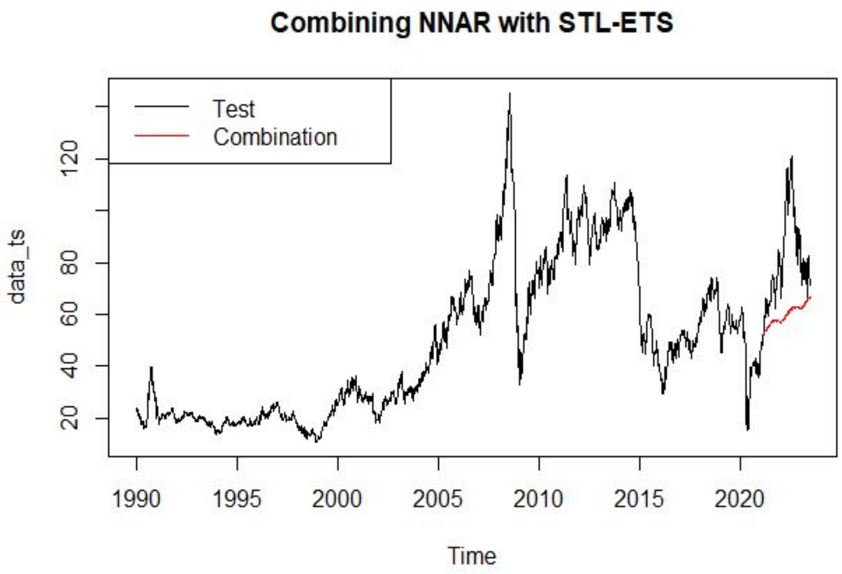
Observation: The simple average Combination model is unable to beat NNAR, motivating custom meta models.



Meta Model 1: simple average

Motivation: The optimal NNAR model has captured non-linear trend and magnitude but is unable to capture seasonality. Hence, we combine our previous STL-ETS model with NNAR using simple average.

Observation: The RMSE is lower than the previous combination model but still higher than the vanilla NNAR model.

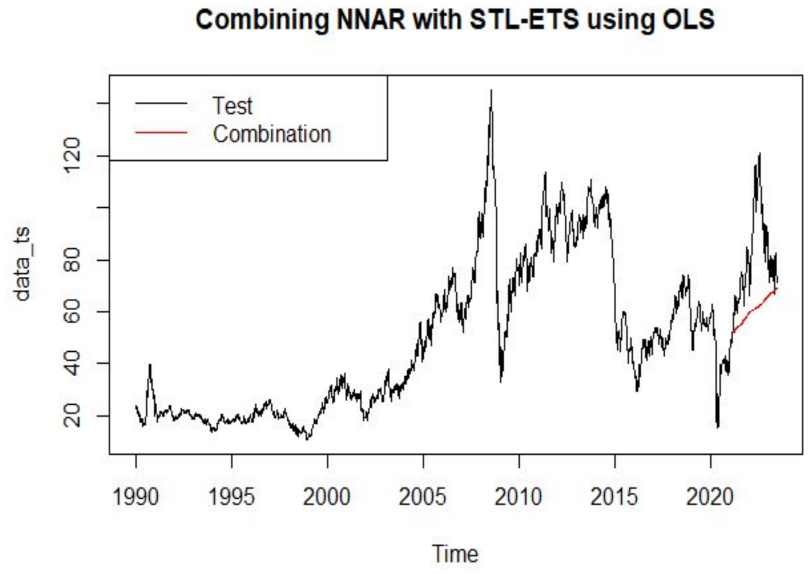


RMSE: 26.03

Meta Model II: OLS Regression

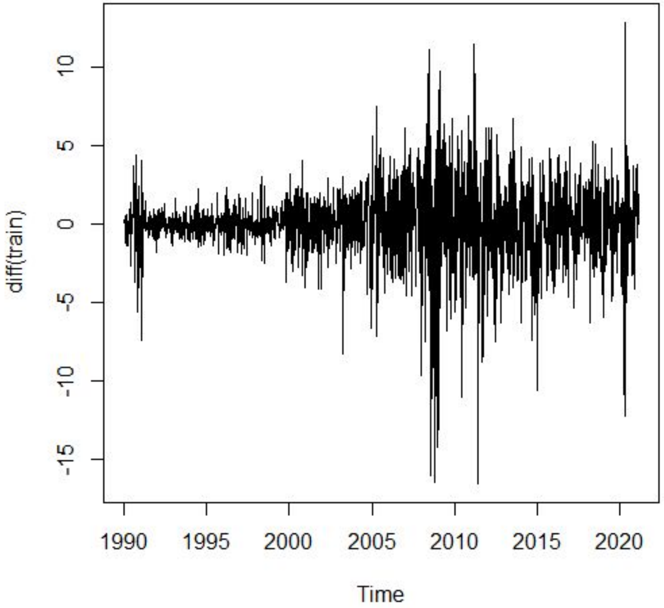
Model (Regressor)	Optimal Weight (OLS beta coefficient)
STL-ETS	-0.17
NNAR (11,1,6)	1.17

Observation: Using OLS regression to find optimal weights for the combination shows that much higher weight is placed on NNAR. This confirms that seasonality is weak. Again, the RMSE here is lower than the previous combination that weighted both models equally. However, the solo NNAR model still performs best.

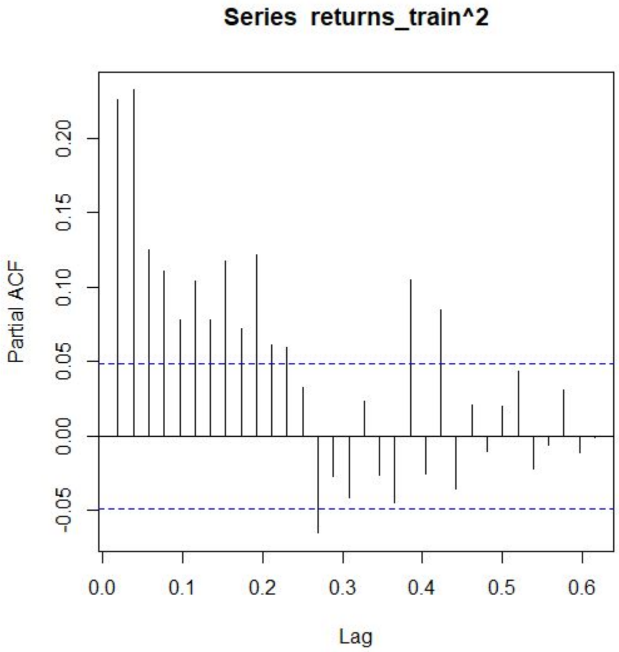


RMSE: 25.2

Conditional variability modeling



Mean Stationarity in Returns.
No Variance/Covariance Stationarity



Persistence is shown in the squared returns

ARCH/GARCH MODEL SELECTION

Volatility Model Garch(1,1), No Mean Model

```
Akaike       4.094672
Bayes        4.111309
Shibata      4.094653
Hannan-Quinn 4.100846
```

Model Description

Alpha of 0.11 shows that the returns of crude oil weighs shocks in oil lightly.

Beta of 0.88 shows that 88% of the past volatility of returns of crude oil is transferred to the current volatility.

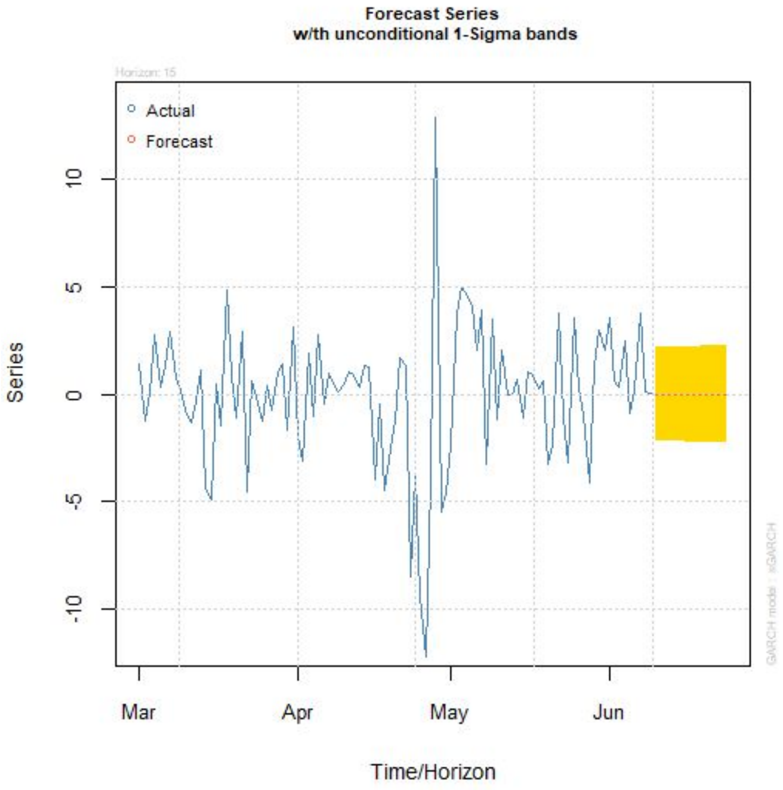
Persistence = α/(1-β) = 0.99

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : sstd

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
omega   0.019053    0.008297   2.2965 0.021649
alpha1  0.110150    0.016652   6.6148 0.000000
beta1   0.888849    0.016318  54.4708 0.000000
skew    0.877349    0.029206  30.0403 0.000000
shape   8.834523    1.614346   5.4725 0.000000
```

Arch/Garch modeling

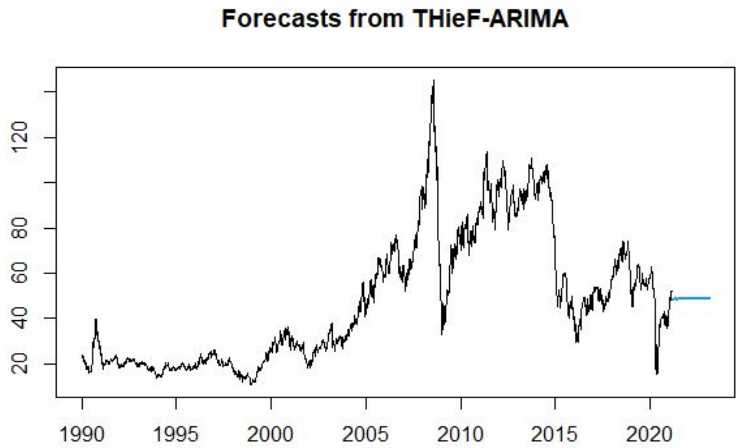
Observations: The model has a mean of zero, inaccurate given the test set, and the standard deviation (sigma) increases by ~0.003 every period. The forecast is too general to be actionable.

σ_{test} = 2.482877 $\sigma_{predicted}$ = 2.223733

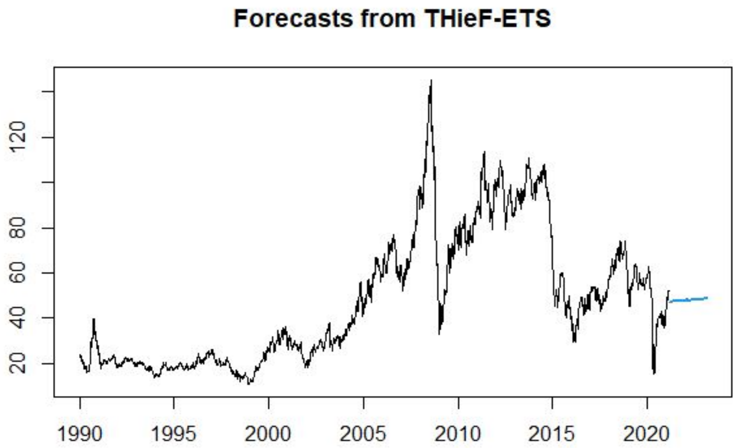


Model For Crude Oil Returns Forecasted Via GARCH

Additional Models: THIEF



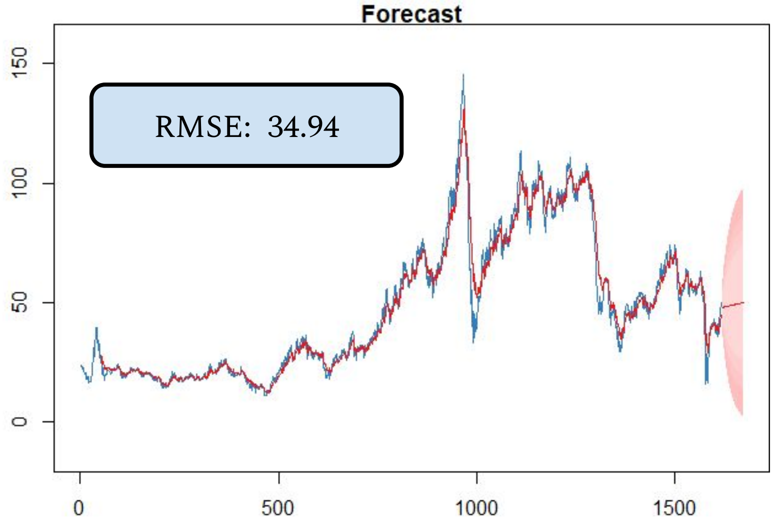
RMSE: 37.8

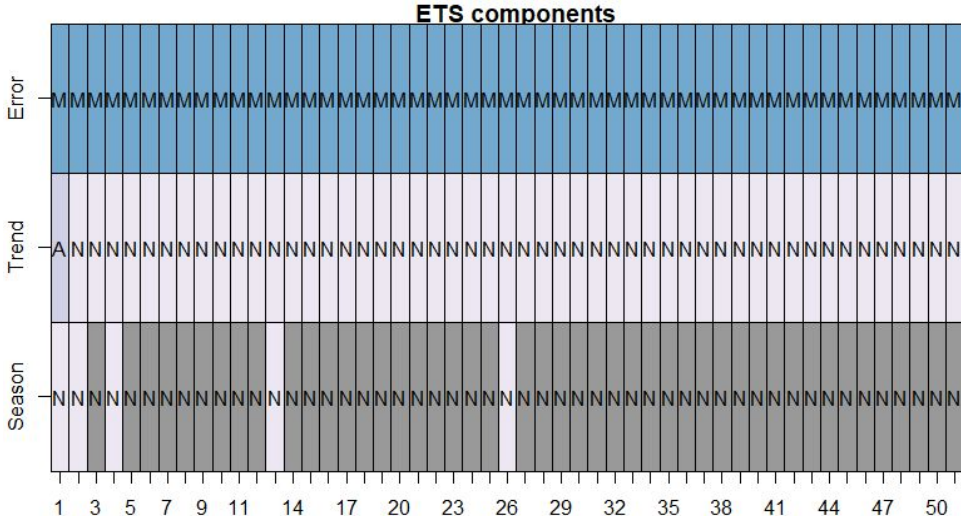


RMSE: 38.27

Model: THieF model hierarchically aggregates the series at different levels. We have used the ARIMA and ETS methods of forecasting to capture different seasonal aggregates.
Observation: The RMSEs are large and fail to produce high accuracy forecasts. Again, including ETS and seasonality worsens the predictive power.

Additional Models: MAPA

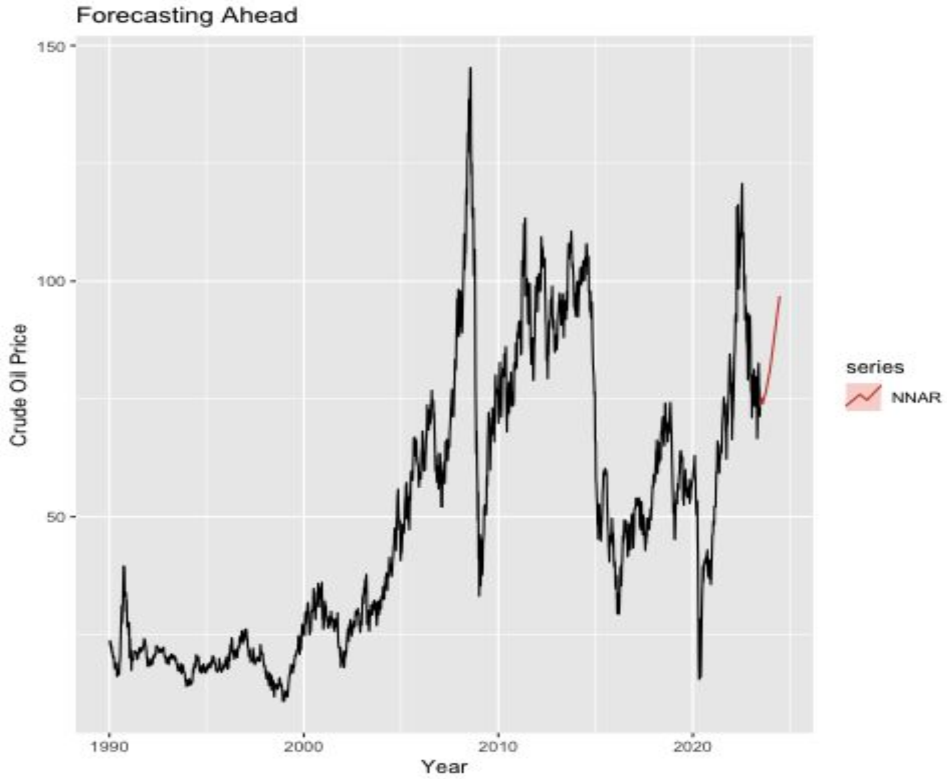




Model: The MAPA model also hierarchically aggregates the series at different levels. It specifically breaks the series into multiple levels and finds unique ETS combinations (as shown above).
Observation: Although the RMSE is lower than THieF, it is still large and produces a relatively flat forecast. Further, the model uses MAN at level 1 and MNN thereafter. Thus, only multiplicative error components are used, motivativing volatility modeling and confirming lack of linear trend and seasonality.

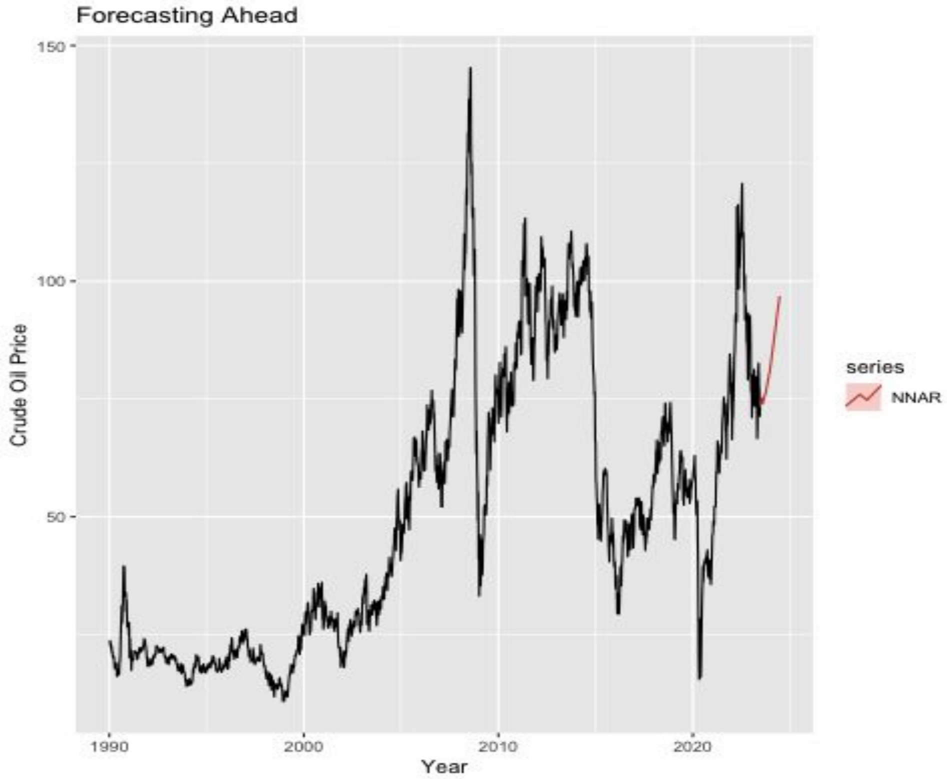
Conclusion and Future Work

Forecasting: A year Ahead



Rise in Oil Prices! :(

Forecasting: A year Ahead



Forecast discussion:

This forecast is for beyond our test set, so it will be interesting to see how consistent the model is with reality moving forward. Just on simple inspection the forecast does seem to capture well past patterns and sudden spikes, especially compared to other models with is a good sign.

Conclusion and future work

Most successful models:

- NNAR
- Seasonality too weak to truly leverage seasonal methods

Future Work: How do we improve the forecast?

- Univariate forecasting is too limited
- Incorporate more features for a data rich neural net approach
- Combine volatility modeling with best model

Would we implement a trading strategy based on our forecasts?

- Not with our own money!

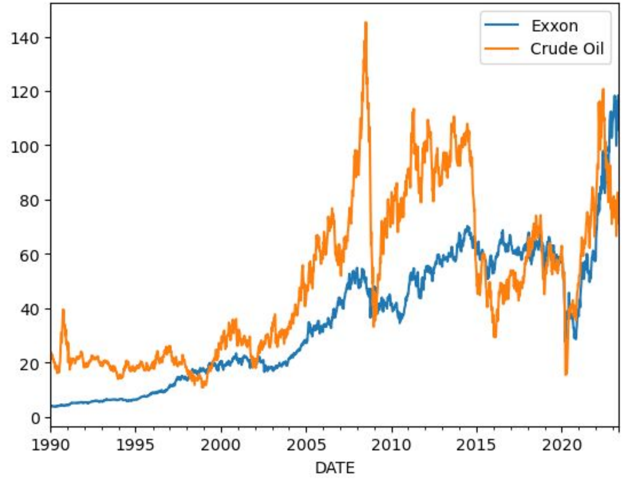
Conclusion and future work

Evaluating the project success and our key takeaways:

Part of our initial motivation was to test the limits of univariate forecasting. We went into our analysis expecting inconsistent model performance, but were actually relatively surprised on how well several of the models performed (especially the neural net). While clearly an imperfect model, this was a satisfying foray into exploring time series that don't behave cleanly, which we expect will be the most valuable forecasts in our future work both academic and professional.

Appendix

VAR Model



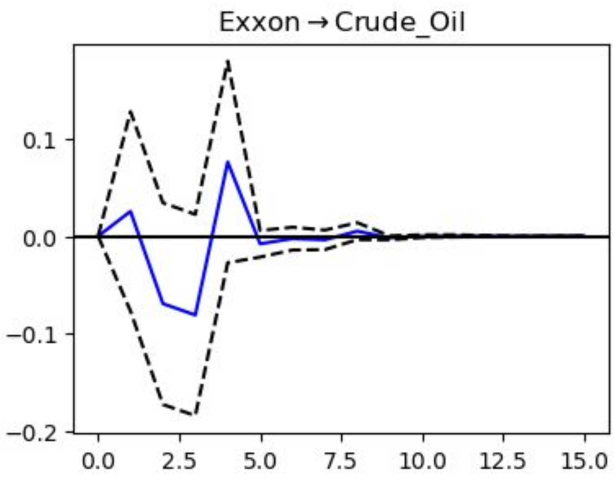
Results for equation Crude_Oil

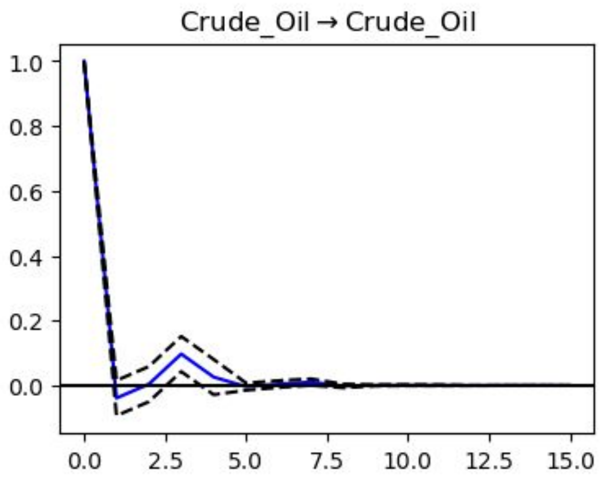
	coefficient	std. error	t-stat	prob
const	0.029677	0.067610	0.439	0.661
L1.Exxon	0.025735	0.052785	0.488	0.626
L1.Crude_Oil	-0.038948	0.027691	-1.407	0.160
L2.Exxon	-0.068139	0.053425	-1.275	0.202
L2.Crude_Oil	0.002951	0.027681	0.107	0.915
L3.Exxon	-0.087742	0.053388	-1.643	0.100
L3.Crude_Oil	0.096974	0.027719	3.498	0.000
L4.Exxon	0.073170	0.053442	1.369	0.171
L4.Crude_Oil	0.028247	0.027878	1.013	0.311

Model: Testing if Exxon stock prices can be used to predict oil prices using VAR models.

Results: The above VAR model summary shows that none of the Exxon lags have significant predictive power for oil prices. Only the third week lagged oil price impacts current oil prices.

VAR Model





Granger causality F-test. H_0: Exxon does not Granger-cause Crude_Oil. Conclusion: fail to reject H_0 at 5% significance level.

Test statistic	Critical value	p-value	df
1.542	2.375	0.187	(4, 3454)

Results: The Impulse Response function graphs along with the Granger Causality test show that Exxon prices do not cause Crude Oil prices. A shock in Exxon prices does not have a discernible effect on Oil prices. Hence, VAR model was abandoned.

References

Data Sources:
FRED Crude Oil Prices West Texas Intermediate (WTI)
Yahoo Finance (Exxon Stock Price)

Cartoon Source I: Peixeiro, M. (2023, April 2). *The Complete Guide to Time Series Analysis and forecasting*. Medium. https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775

Cartoon Source II: *Hedgeye Risk Management: Cartoon of the day: Crude reality*. Hedgeye. (2015, January 9). https://app.hedgeye.com/insights/41657-cartoon-of-the-day-crude-reality