Historical Data Analysis & Price Prediction Models

Petrochemical Index (ICIS), Brent (US EIA)

Ng Qi Xuan

Machine Learning Intern

Digital Team

Project Scope

Forecast Raw Material Pricing & Industry Melt till Dec 2025

Forecast Petrochemical Index till Dec 2025

Machine learning (ML) modelling

Extract, transform and load data to

Amazon SageMaker

Data Inputs: ICIS and raw_mat from 2021 to 2023

month	ICIS_China_high	ICIS_China_low	ICIS_Korea_high	ICIS_Korea_low	ICIS_SEA_high	ICIS_SEA_low
Jan 2021						
Feb 2021						
. . . .						
Oct 2023						
Nov 2023						
Dec 2023						

month	rm_pta	rm_meg	rm_indmelt
Jan 2021			
Feb 2021			
. . . .			
Oct 2023			
Nov 2023			
Dec 2023			

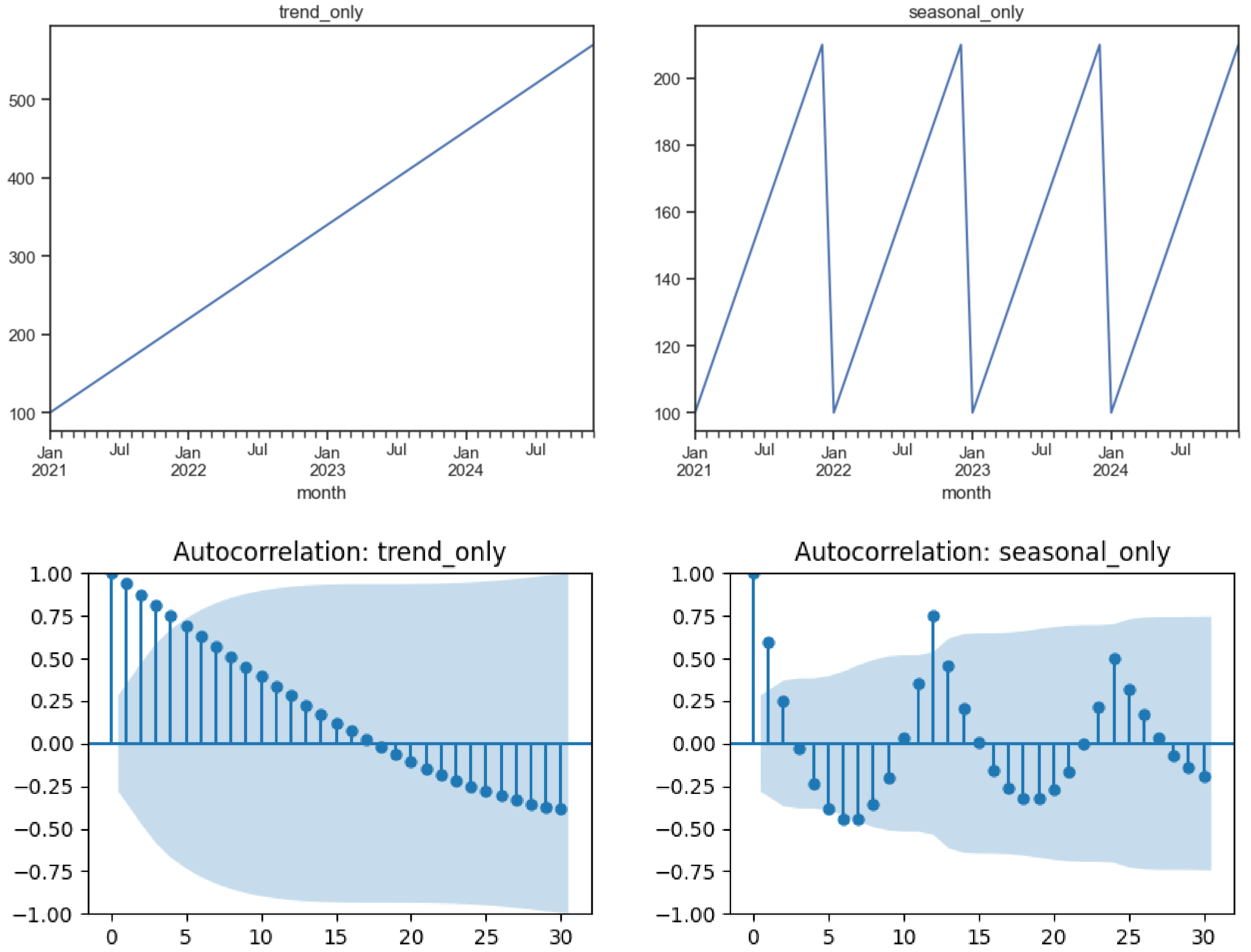
rm_indmelt = 0.84 × rm_pta + 0.34 × rm_meg

Data Visualization – Identifying underlying trends and seasonal patterns with autocorrelation plots

- Autocorrelation, *f(x)* measures the correlation between the price at a particular month and *x* month ago.

- Seasonal pattern: repetitive price fluctuations at fixed time period

- Strong peak at time lag = 12 for seasonal_only suggests that the seasonal pattern repeats every 12 months

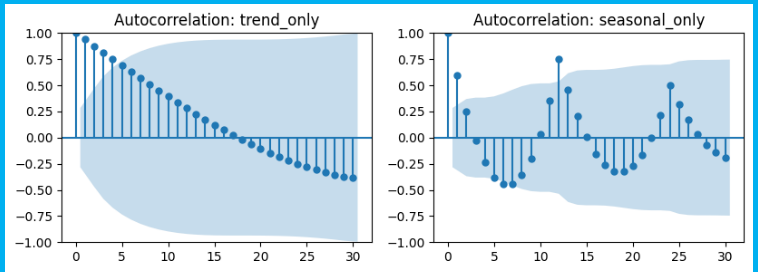


Data Visualization – Identifying underlying trends and seasonal patterns with autocorrelation plots

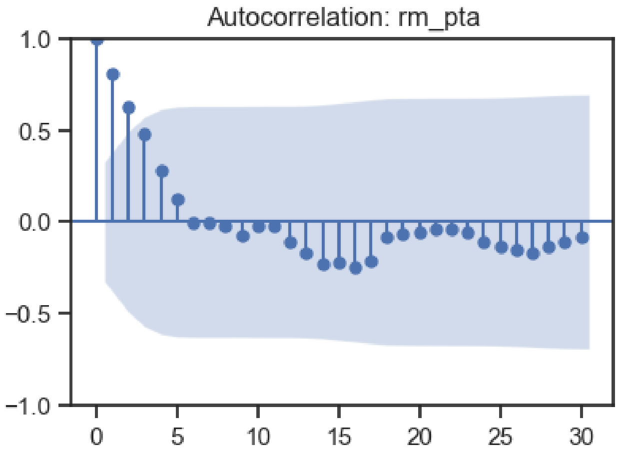
- Autocorrelation, *f(x)* measures the correlation between the price at a particular month and *x* month ago.

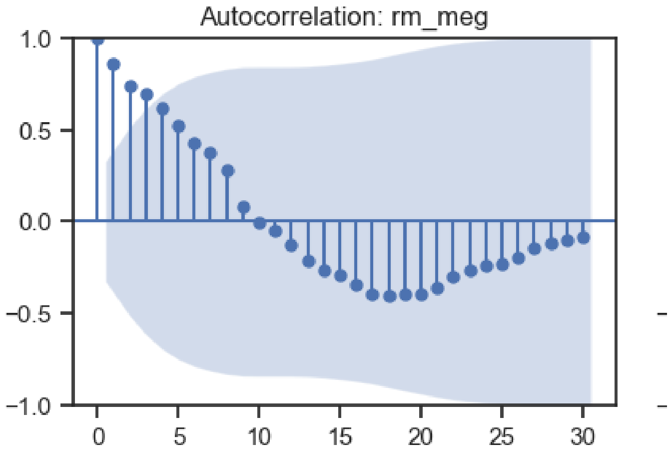
- Seasonal pattern: repetitive price fluctuations at fixed time period

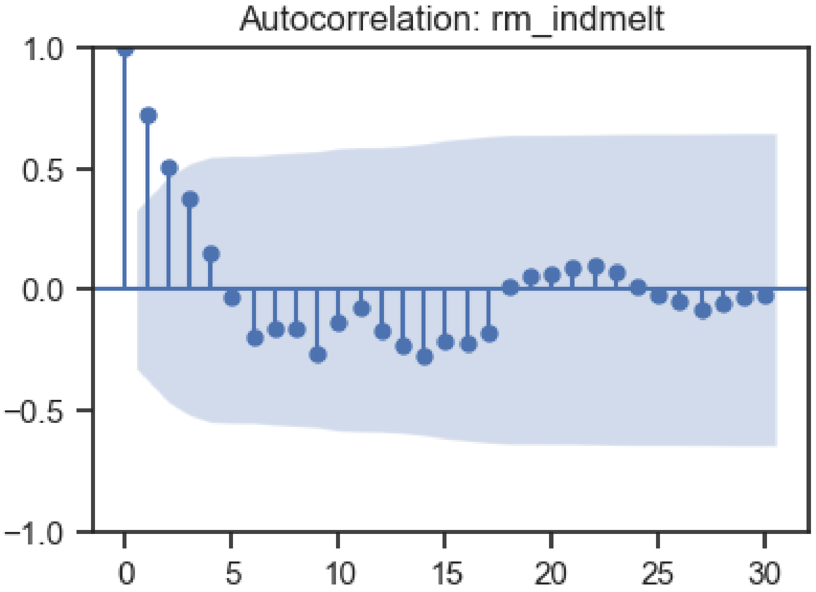
- rm_indmelt has both trend and seasonal components



Graph reference for trend_only time series and seasonal_only time series







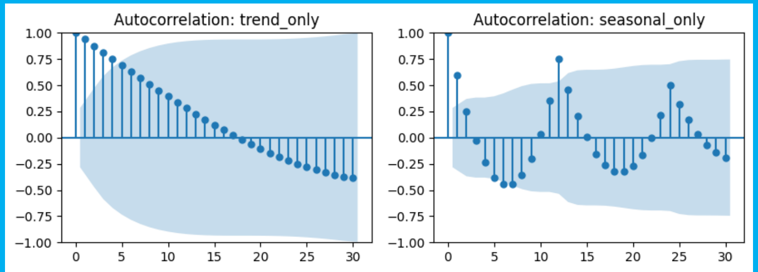
Our vision: To be a world-class sustainable chemical company making great products for society.

<u>Data</u> Visualization – Identifying underlying trends and seasonal patterns with autocorrelation plots

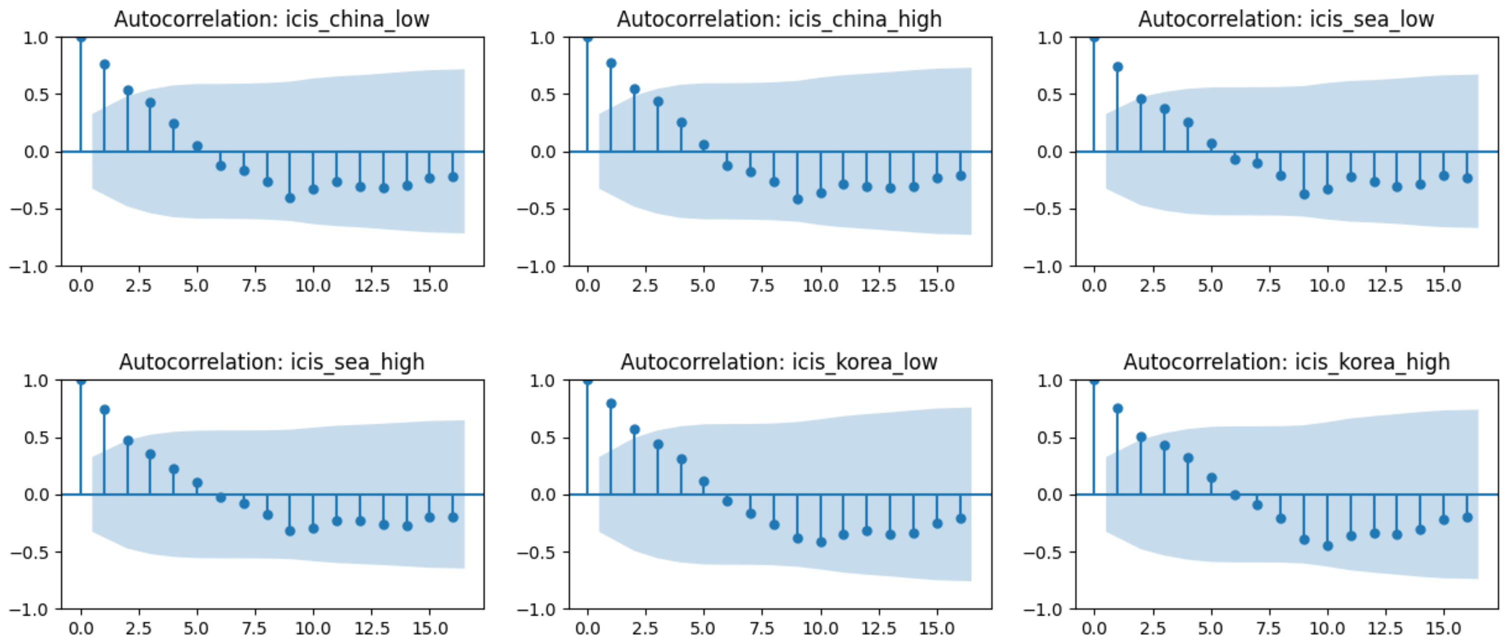
- Autocorrelation, *f(x)* measures the correlation between the price at a particular month and *x* month ago.

- Seasonal pattern: repetitive price fluctuations at fixed time period

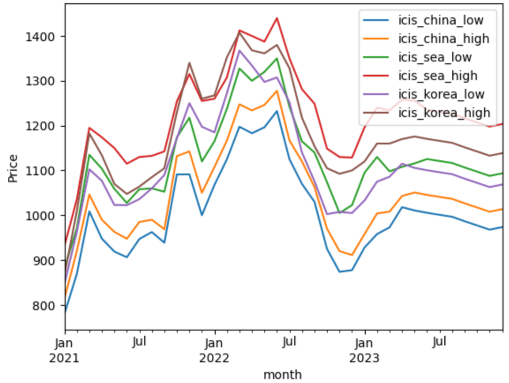
- ICIS displays strong trend and a mild seasonal pattern



Graph reference for trend_only time series and seasonal_only time series



Similar autocorrelation plots expected since these indices exhibit strong correlation among each other.



Our vision: To be a world-class sustainable chemical company making great products for society.

INDORAMA VENTURES

Data Visualization – Time Series Decomposition

Separates a time series into its trend, seasonality, and residual, in order to better understand and analyze the underlying patterns and variations. This decomposition allows more accurate forecasting by identifying key features in the time series.

month	Y
Jan 2021	
Feb 2021	
. . . .	
Oct 2023	
Nov 2023	
Dec 2023	

decomposition →

month	Y_trend	Y_seasonal	Y_residual
Jan 2021			
Feb 2021			
. . . .			
Oct 2023			
Nov 2023			
Dec 2023			

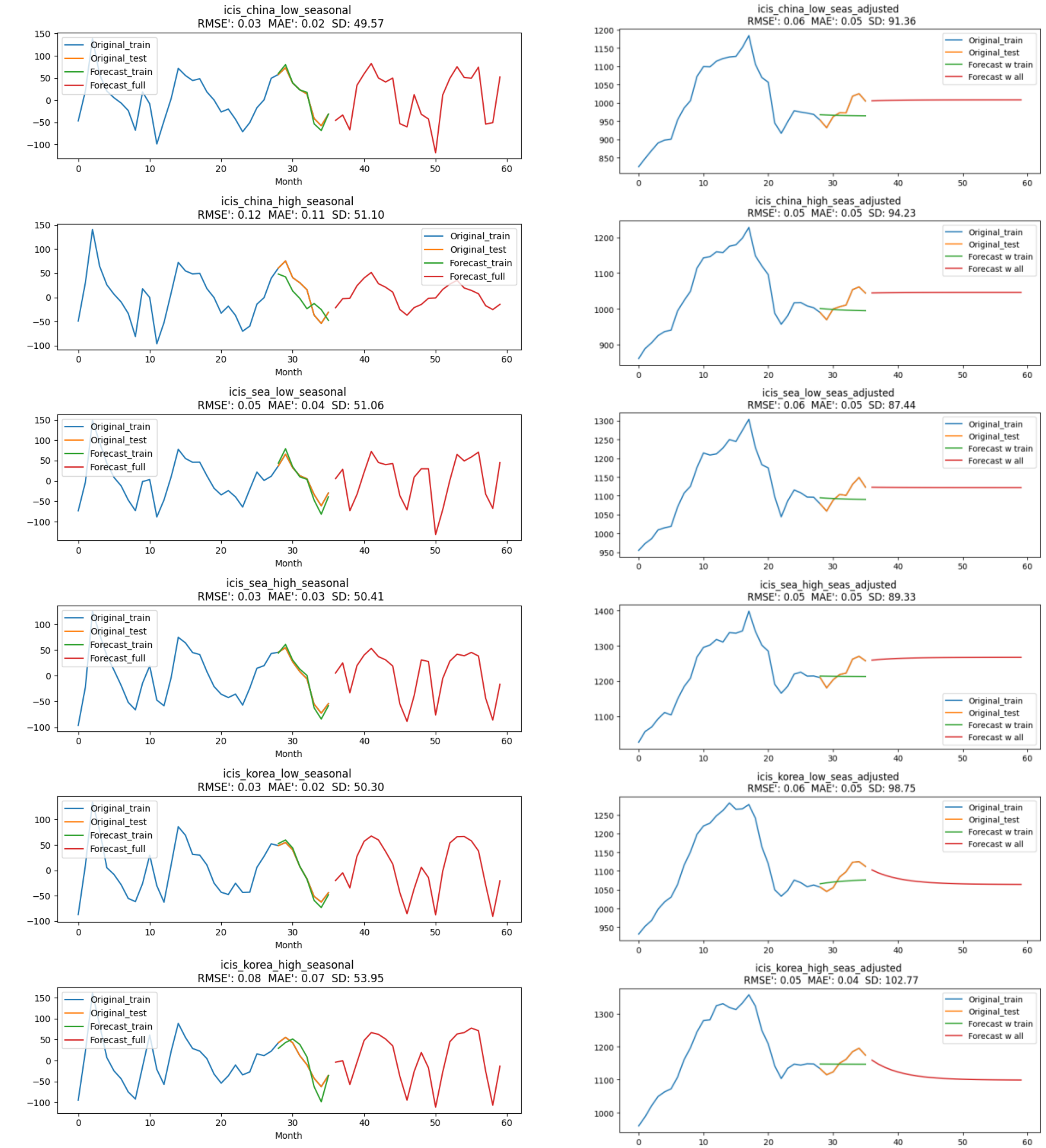
Y = Y_trend + Y_seasonal + Y_residual

Initial Results

- Y_seas_adjusted = Y - Y_seasonal

- Modelling performance evaluated by RMSE’ and MAE’
  - RMSE: Root Mean Square Error
  - RMSE': RMSE/(test_data(max)-test_data(min))
  - MAE: Mean Average Error
  - MAE': MAE/(test_data(max)-test_data(min))

- The models perform better on the seasonal components compared to the seasonally adjusted component.



Challenges

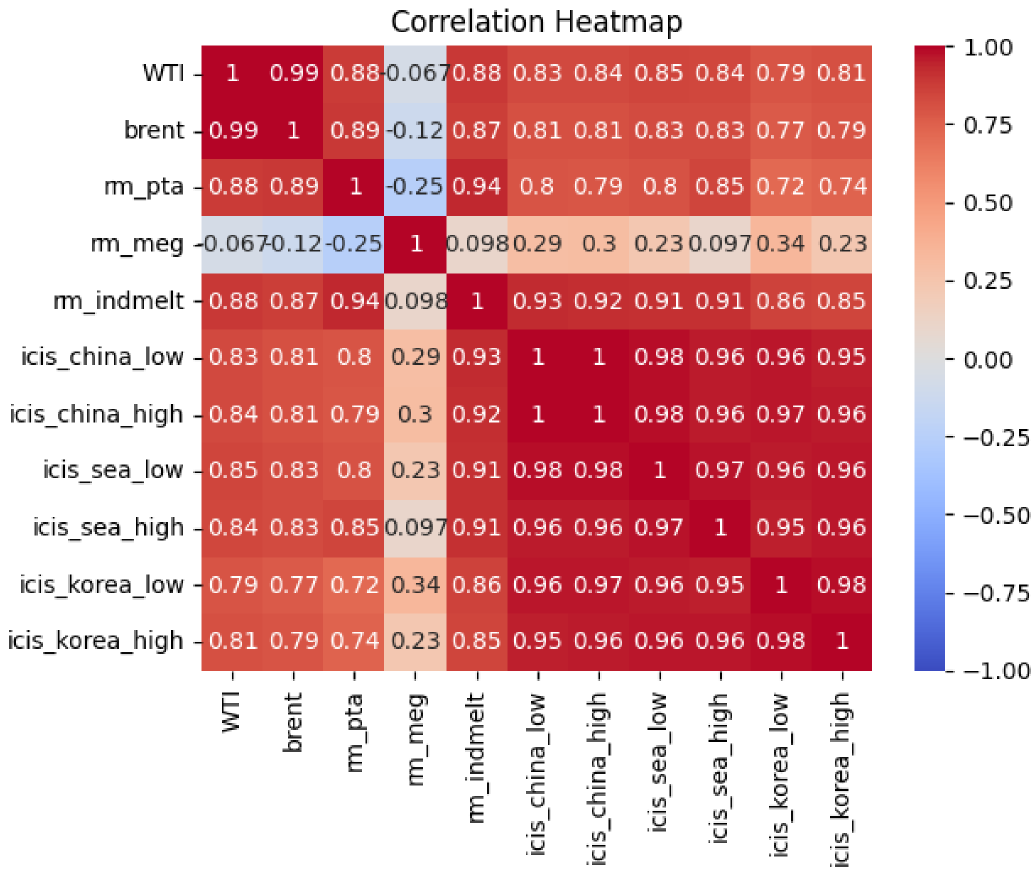
- High risk of overfitting due to insufficient input data. Currently we only have 36 monthly values to forecast the next 24 values.
- In light of current geopolitical uncertainty, the modelling assumption that all else remains constant will not hold true.

Solution

1. Identify a common variable that
   a) affects both rm_indmelt and ICIS
   b) Has more historical data available
2. Plot heatmap to confirm that this variable has a high correlation with rm_ indmelt and ICIS
3. Use this variable to train time series models for forecasting
4. Perform regression analysis to work backwards and obtain forecasted rm_indmelt and ICIS

Data Visualization – Crude oil

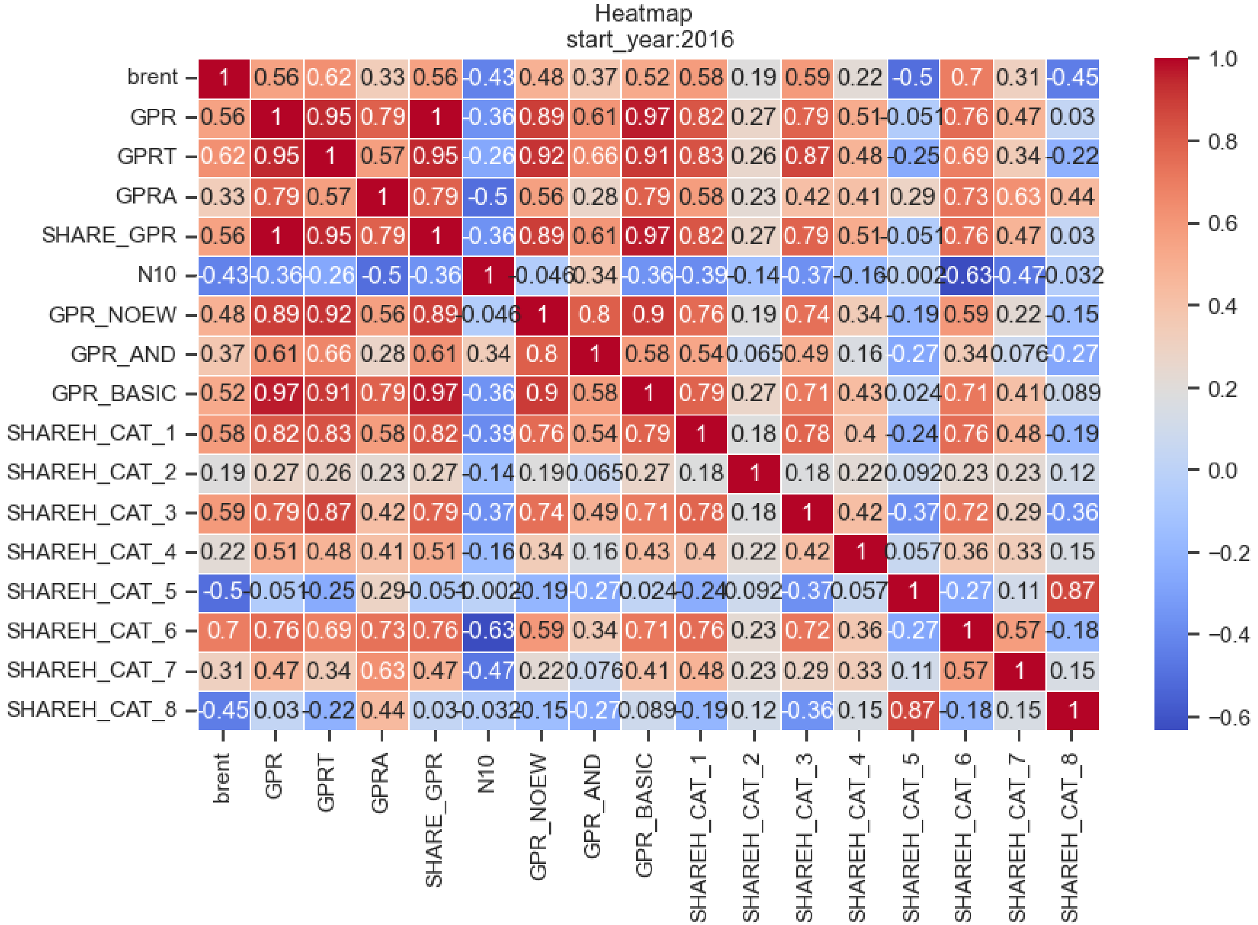
Two types of crude oil, brent and WTI, have readily available historical data from 1987 onwards published by US Energy Information Administration (EIA) online.



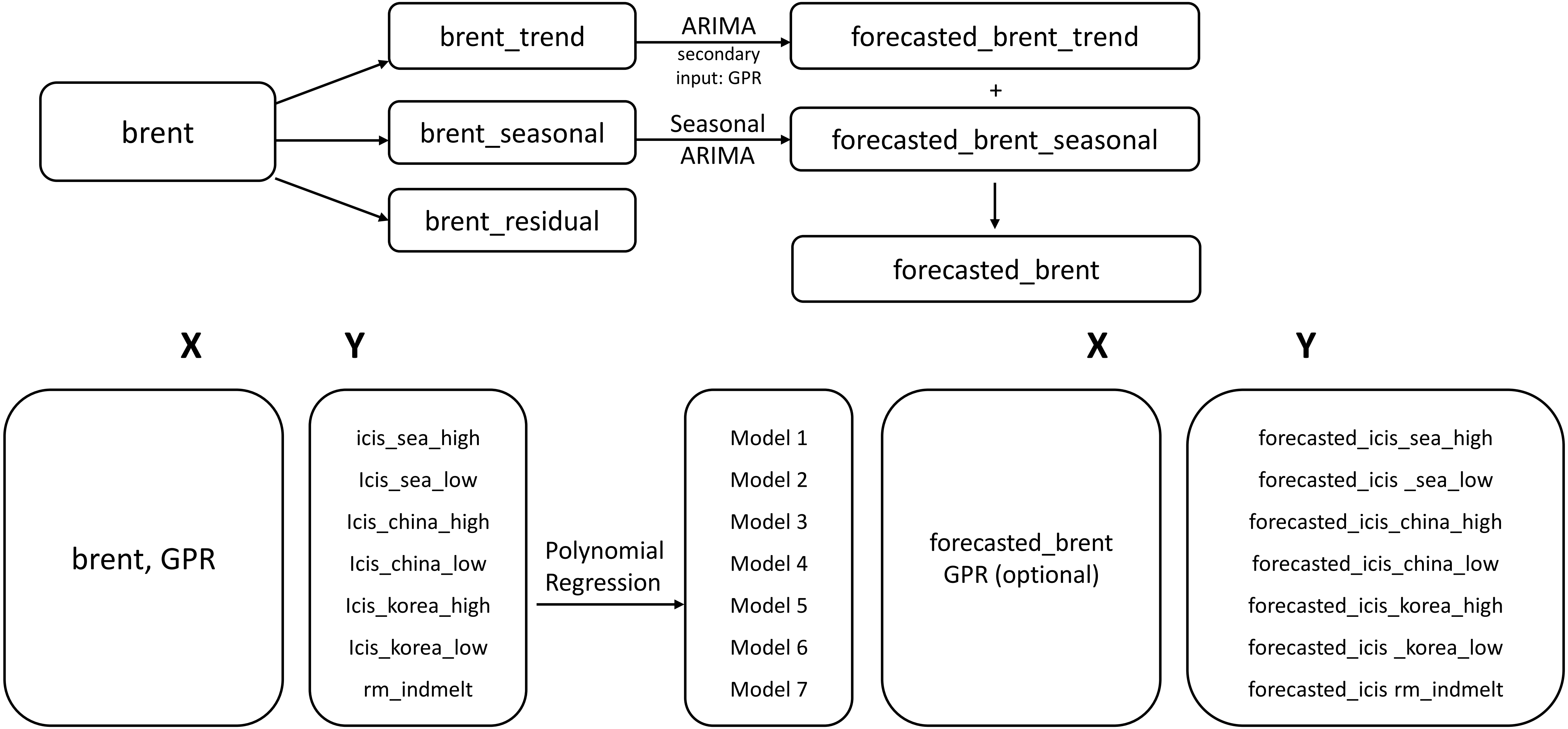
- Crude oil, as a raw material, has strong influence on ICIS and rm_indmelt.
- rm_meg has very low correlation with crude oil. Proceed to directly forecast rm_indmelt instead of its intermediate products.
- WTI and Brent exhibits high correlation to each other, choosing either one is sufficient for time series modelling.

Model Fine Tuning – Additional secondary inputs

Geopolitical risk has been recognized as a significant factor influencing financial and economic variables, and its impact on commodity prices, particularly in the energy sector, is well-documented. Incorporating GPR into the ARIMA model shall capture the potential effects of geopolitical events and uncertainties on the overall trend pricing.

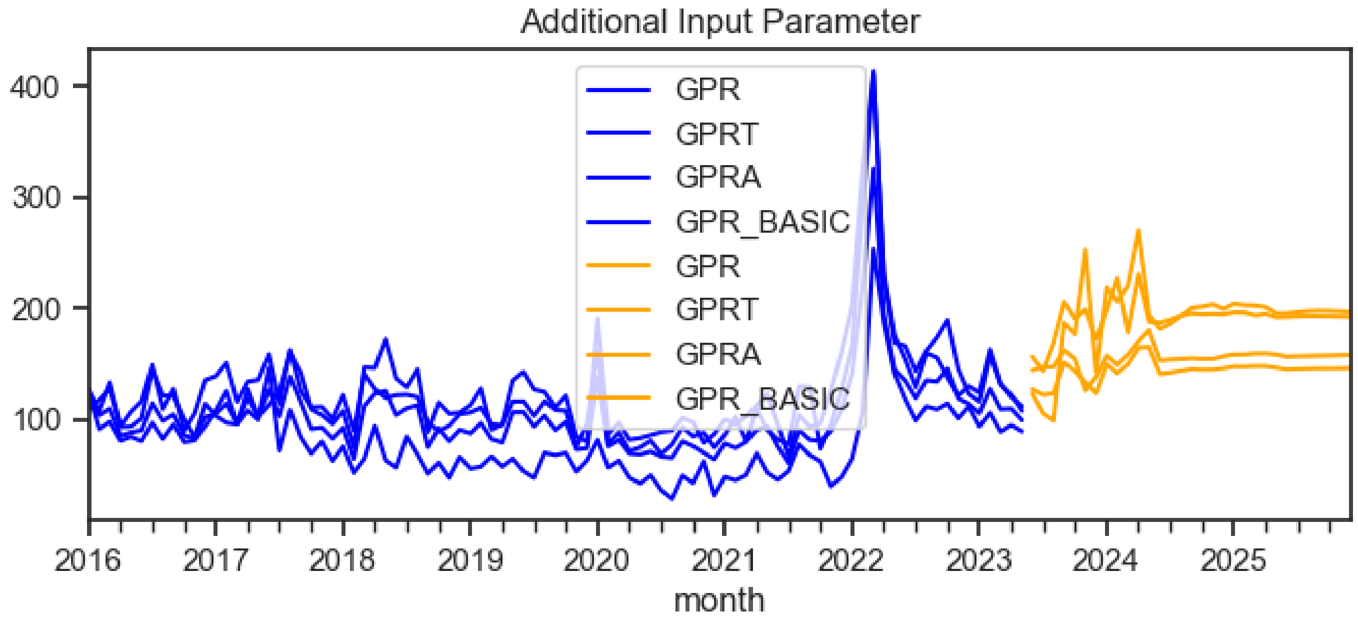


ML Modelling Flowchart



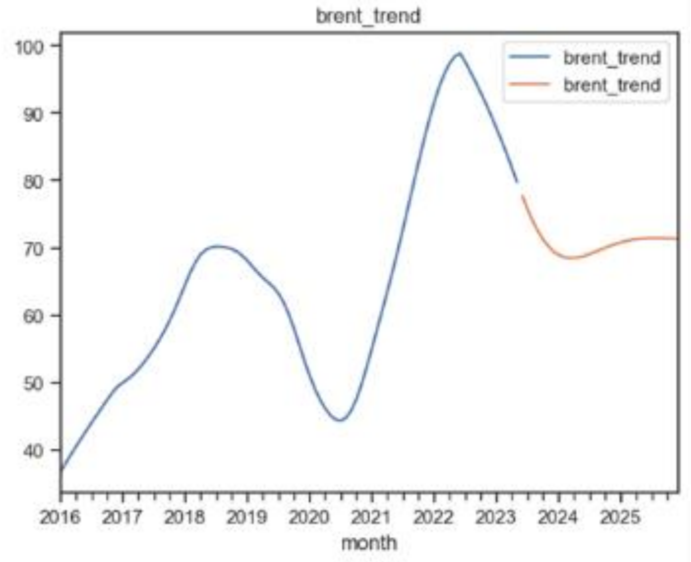
Modelling Assumptions

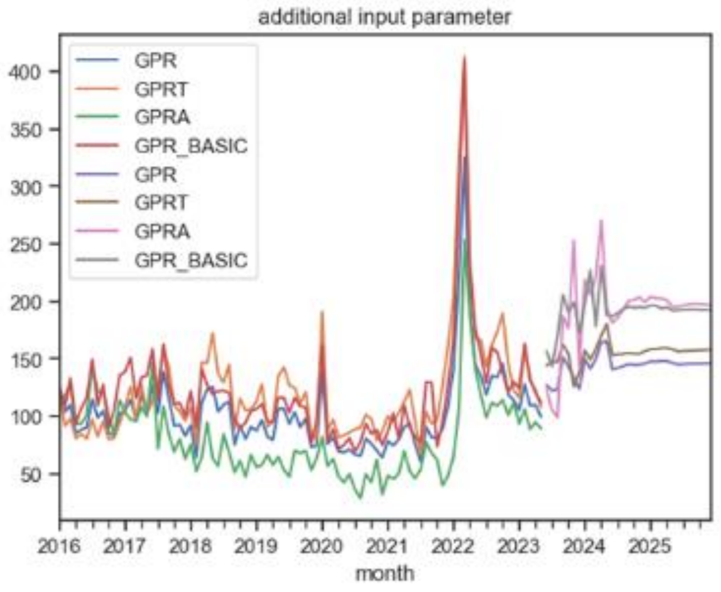
- Past historical events mostly occur independently of each other.
- Created a simple damped exponential smoothing model to 'forecast' future geopolitical risk index (GPR).
- As the time lag increases, the effects of each geopolitical event decrease, hence the forecast eventually approaches a constant.



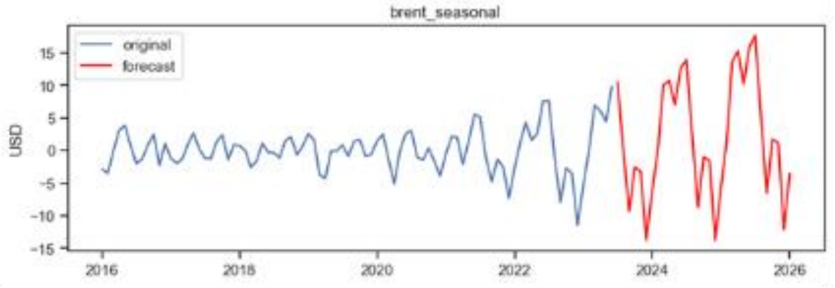
INDORAMA VENTURES

Crude Oil ML Modelling Overview





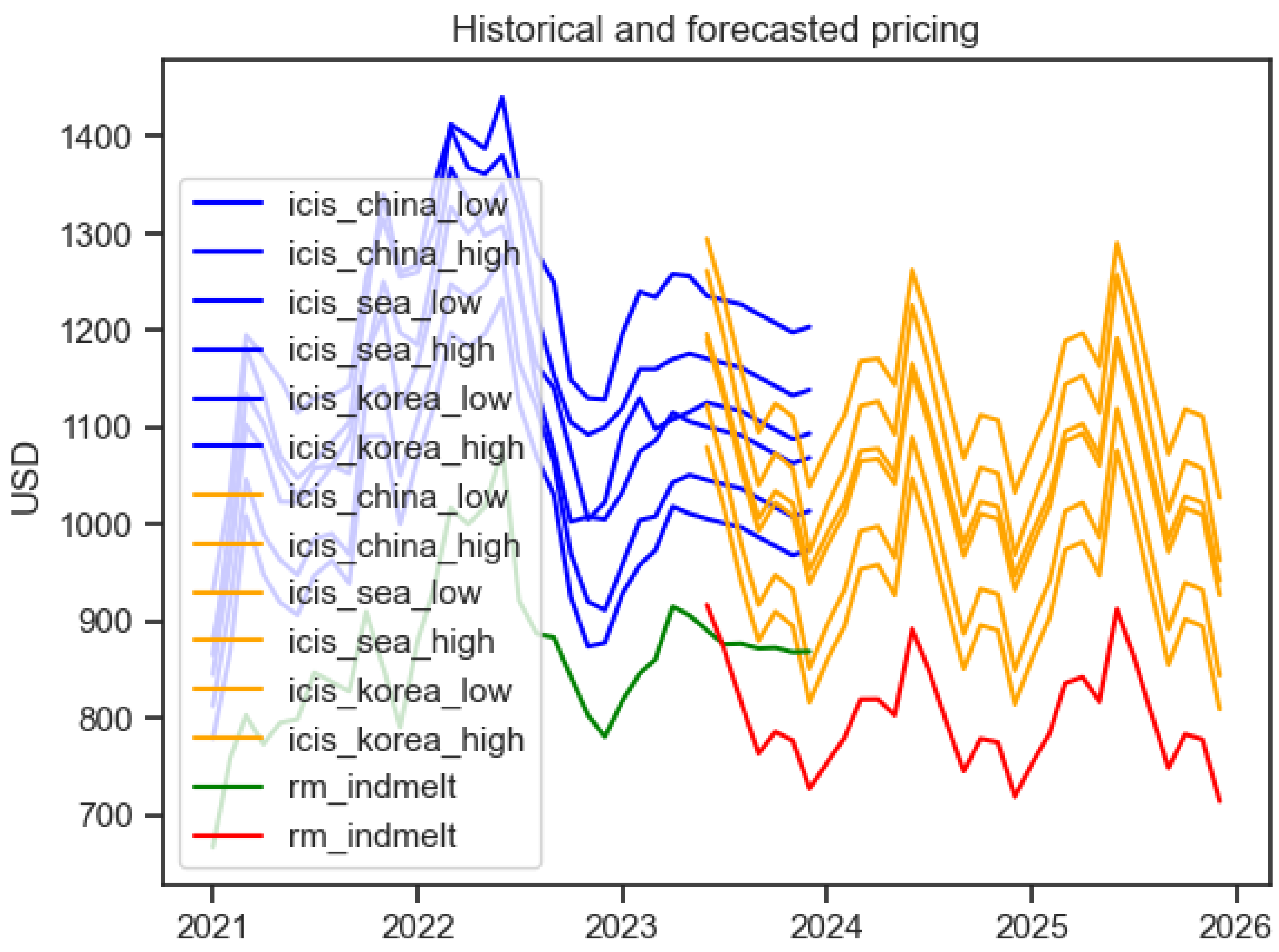
+



=

Our vision: To be a world-class sustainable chemical company making great products for society.

Forecast – ICIS & rm_indmelt

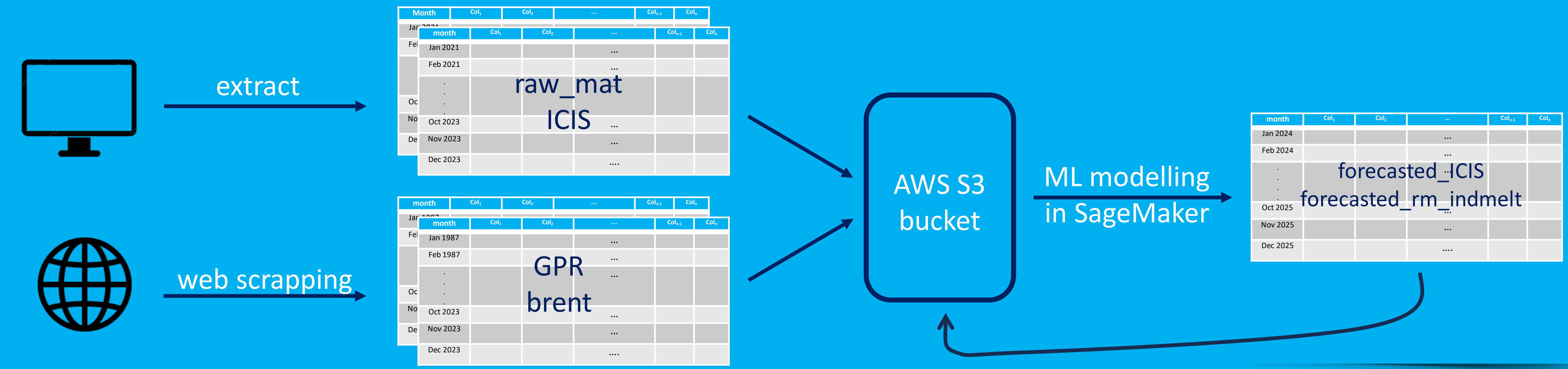


- The overall decreasing trend does not necessarily indicate a decline in crude oil pricing from its initial levels.
- Instead, it suggests that the prices are returning to a more normal level due to supply disruptions, geopolitical tensions, and market uncertainty.
- Users should note that such modelling assumes that no significant geopolitical events emerged in the next few years.

Making a Pipeline in Sagemaker

Compiled the following steps into an .ipynb file

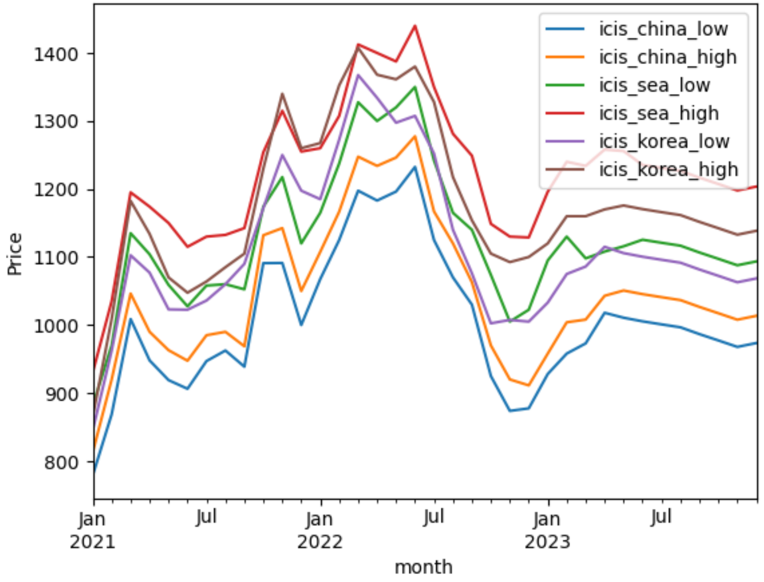
1. Automate data upload into AWS S3 Bucket
2. Data Preprocessing
3. Data Visualization
4. Train ML models
5. Deploy ML Models
6. Store ML predictions in AWS S3 Bucket

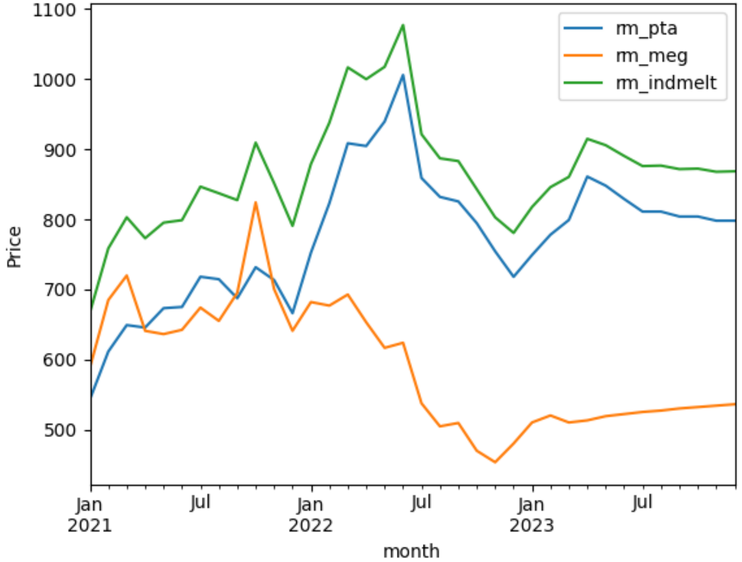


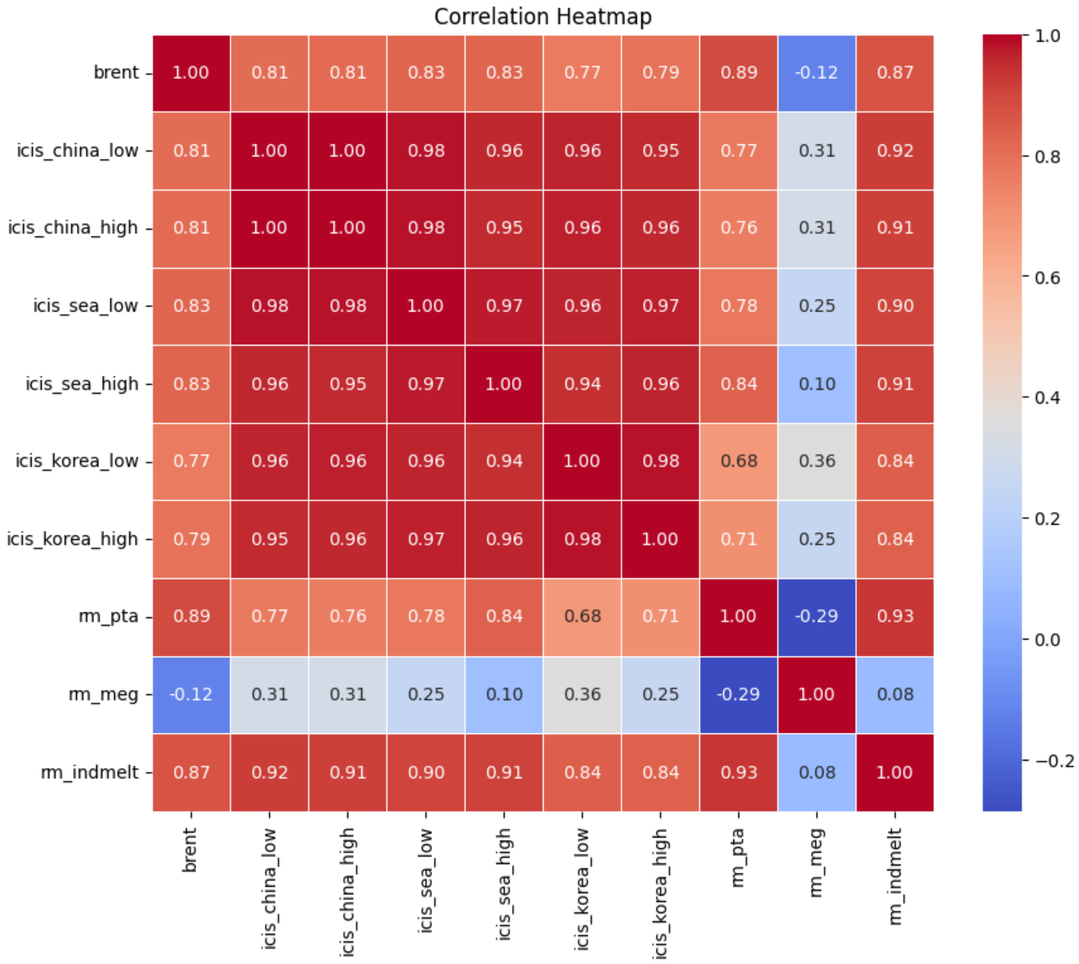
https://www.matteoiacoviello.com/gpr.htm/gpr_files/data_gpr_export.xls
https://www.eia.gov/dnav/pet/

ARCHIVE

Data Visualization – Finding correlation between ICIS and industry melt cost





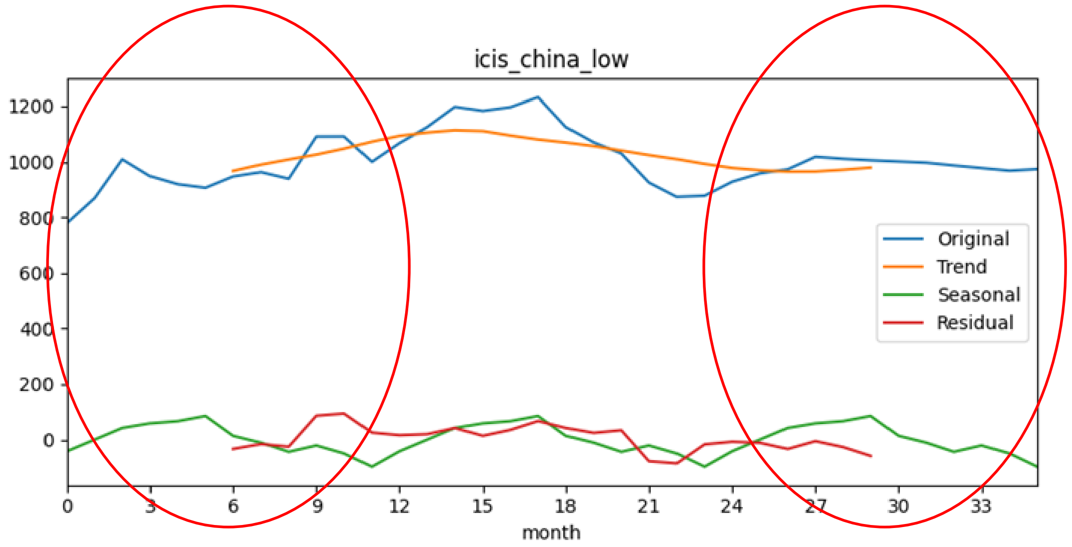


analysis

rm_indmelt = 0.84 × rm_pta + 0.34 × rm_meg

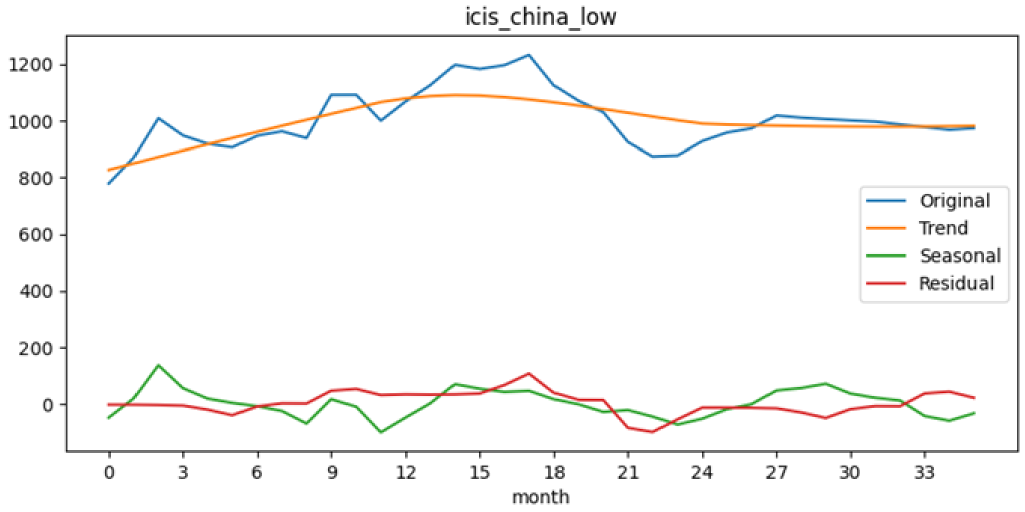
Data Visualization – Time Series Decomposition

Classical Decomposition



- Simpler algorithm
- Unintended missing values at the start and end of time series

STL Decomposition (chosen)



- More complex algorithm
- No missing values at the start and end of the time series

Our vision: To be a world-class sustainable chemical company making great products for society.

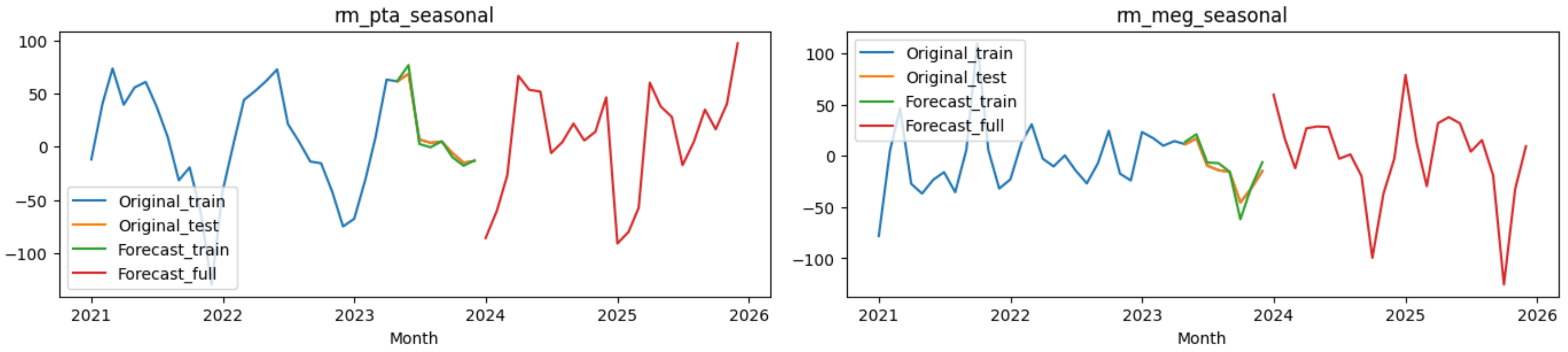
Time Series Modelling Options

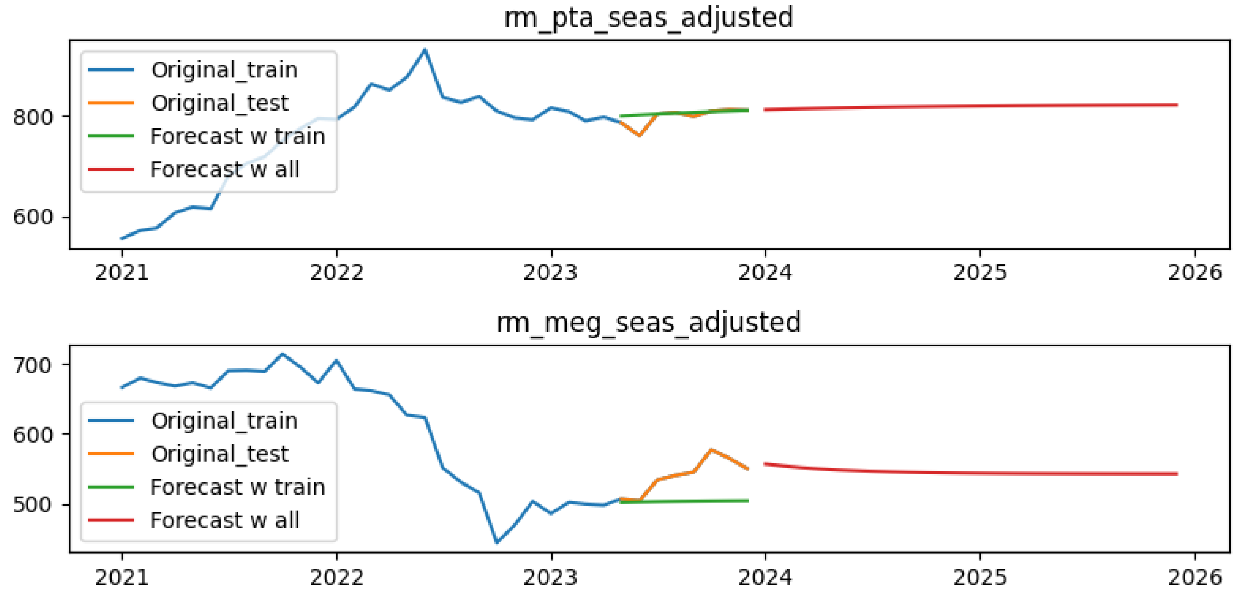
- Exponential Smoothing
  - applies exponentially decreasing weights to past observations
  - forecast is a weighted sum of past observations
  - $\hat{Y}_t = \alpha * Y_{t-1} + (1 - \alpha) * \hat{Y}_{t-1}$
- Holtz Winter Damped Exponential Smoothing
  - Extends the exponential smoothing model with a damping factor to reduce the impact of extreme observations on future forecasts.
- Autoregressive Integrated Moving Average (ARIMA)
  - Autoregressive (AR) component represents the dependence of the current observation on previous observations
  - Differencing (I) component removes the trend and seasonality from the data
  - Moving average (MA) component models the dependence on past forecast errors
- Seasonal ARIMA with Exogenous Variables (SARIMAX)
  - Extends the ARIMA model by incorporating seasonal patterns and exogenous variables then influence the target variables

Initial Results

- Modelling performance evaluated by RMSE' and MAE'
  - RMSE: Root Mean Square Error
  - RMSE': RMSE/(test_data(max)-test_data(min))
  - MAE: Mean Average Error
  - MAE': MAE/(test_data(max)-test_data(min))

- With a train/test split of 28/8, the best ARIMA and exponential smoothing models achieve a RMSE' and MAE' ranging from 0.06 to 0.18 across the 6 data sets in ICIS, and 0.05 to 0.09 across the data sets in raw_mat

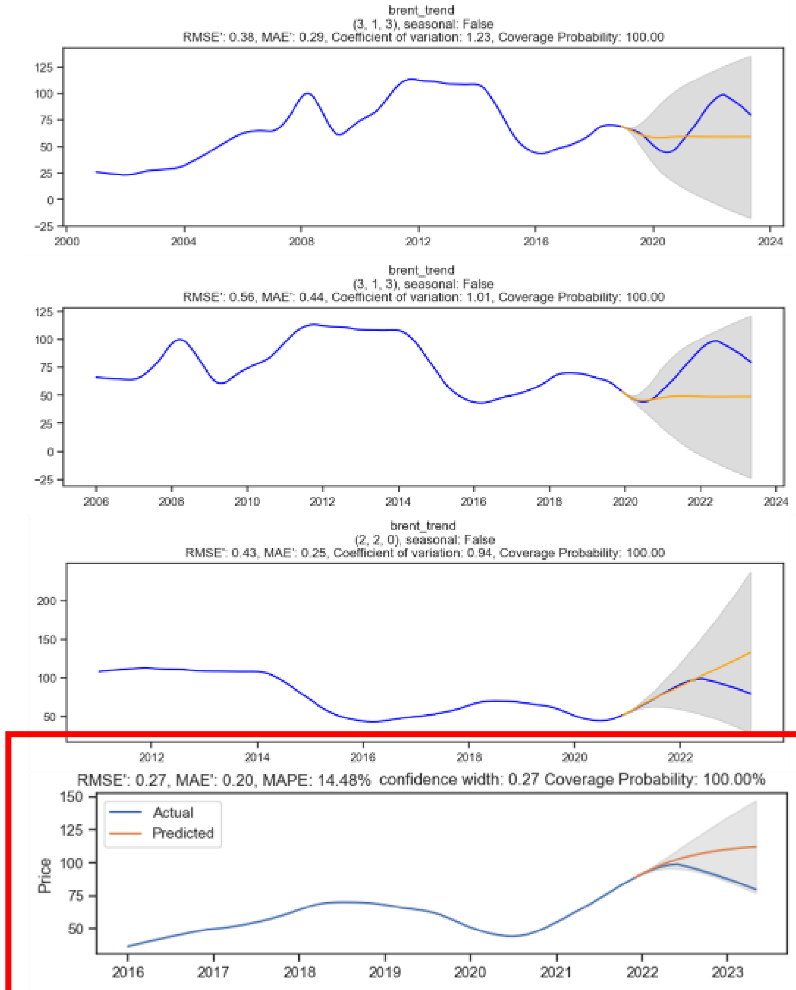




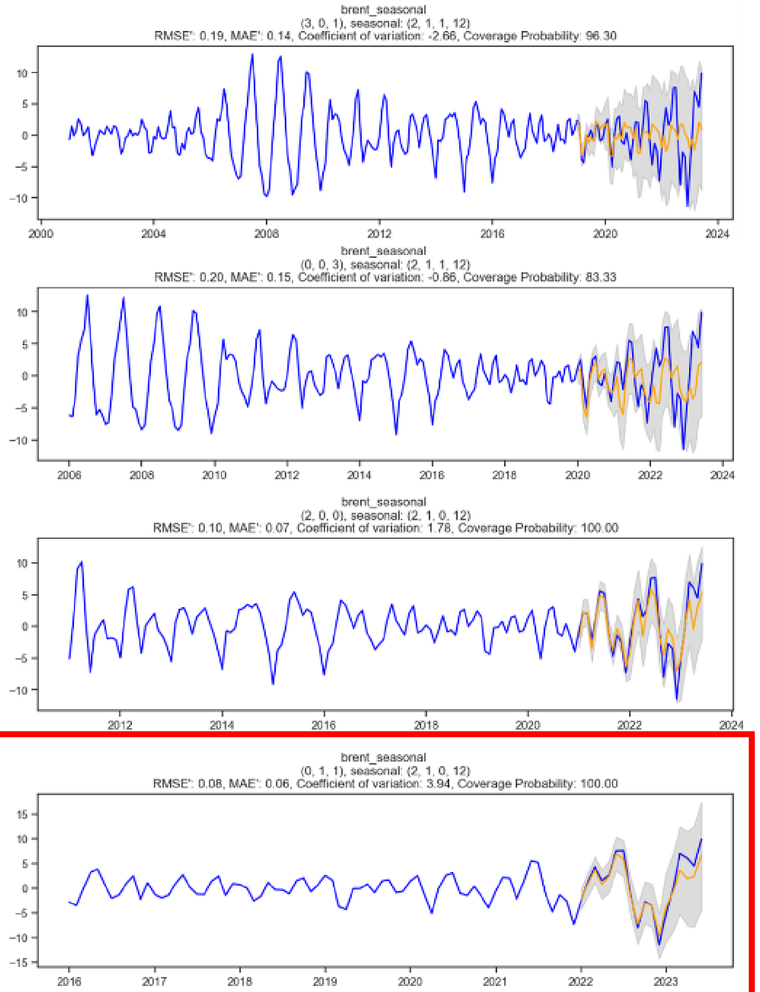
Y_seas_adjusted = Y – Y_seasonal

Model Selection – Timeframe

ARIMA on the trend component



SARIMA on the seasonal component



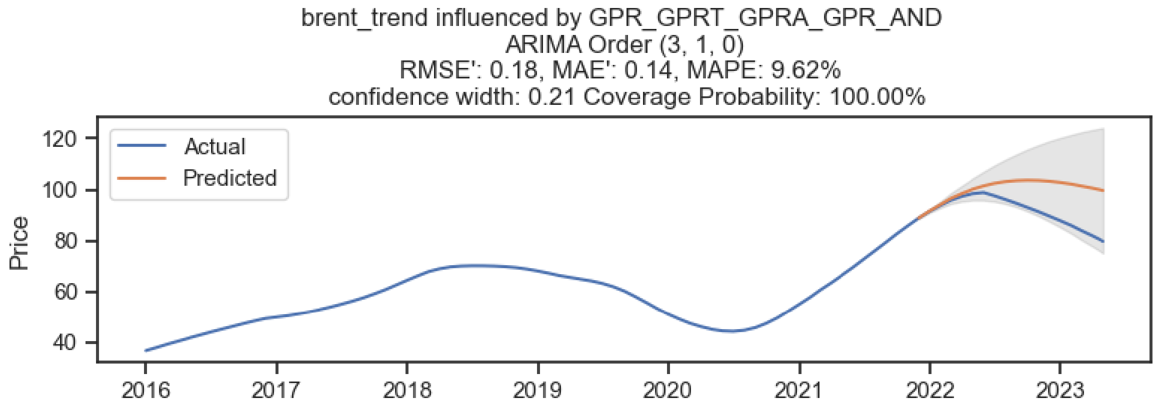
chosen timeframe – 2016 onwards

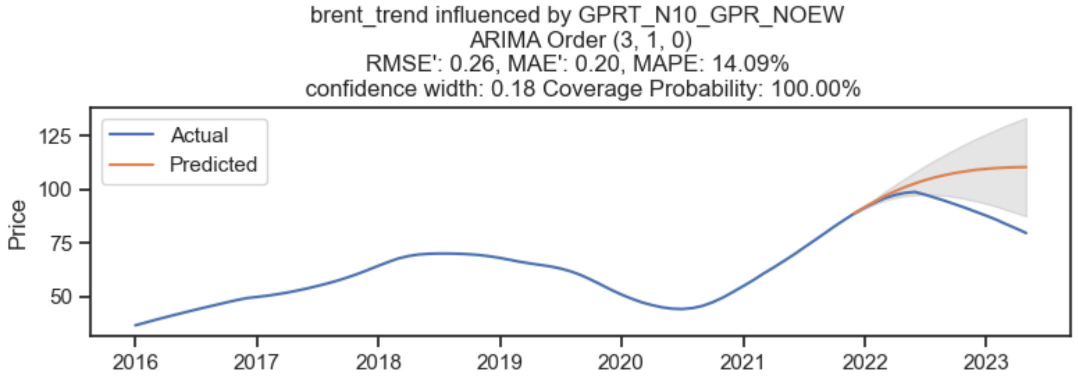
- Models trained with data from 2016 onwards yield the lowest RMSE' and MAE'.
- The trend component has a much wider confidence width denoted in gray, hence the forecast is less reliable.
- Fluctuations in brent_seasonal is only about 10% of brent_trend -> brent_trend has a greater influence on the overall price of brent.
- Subsequent model fining tuning will focus on improving the modelling's accuracy on brent_trend.

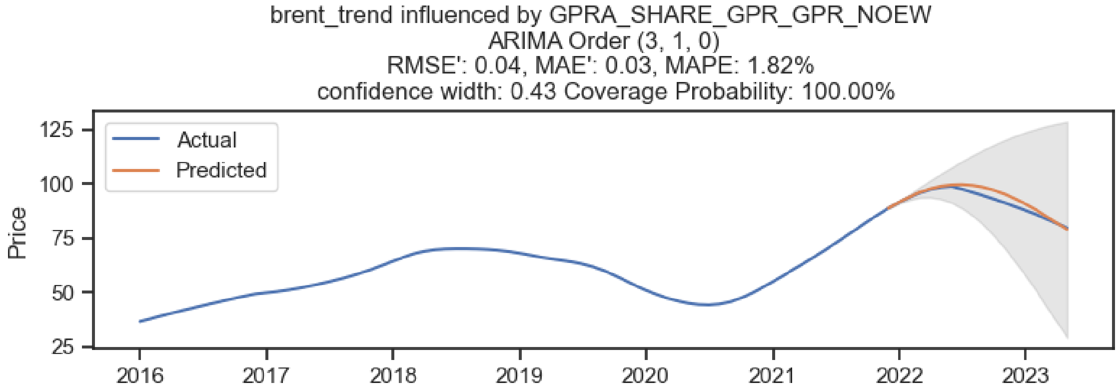
Model Fine Tuning – Additional secondary inputs

The test data performance falls into 3 broad categories:

1) Low error, low confidence, high coverage probability
2) Higher error, higher confidence (i.e. low confidence width), high coverage probability
3) Highest confidence, low coverage probability



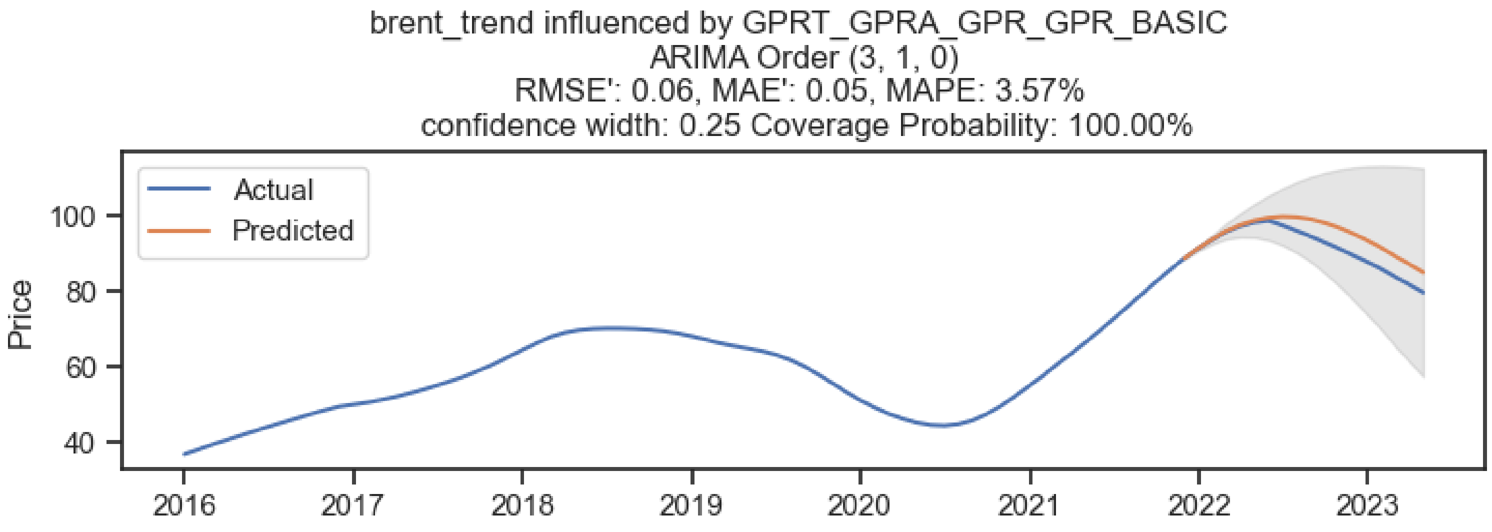




Model Fine Tuning – Additional secondary inputs

We do not want a case where a data point in y_test does not fall within the 99% confidence interval as denoted by the gray shaded area, although a narrower confidence width is preferred.

Hence, we first filter cases where the coverage probability = 100% and the confidence_width < 30%. Then, the exogenous input that give the lowest RMSE’ and MAE’ shall be the best model for brent_trend.



Best model for brent_trend with exogenous inputs