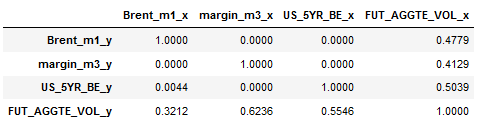
**Analysis**

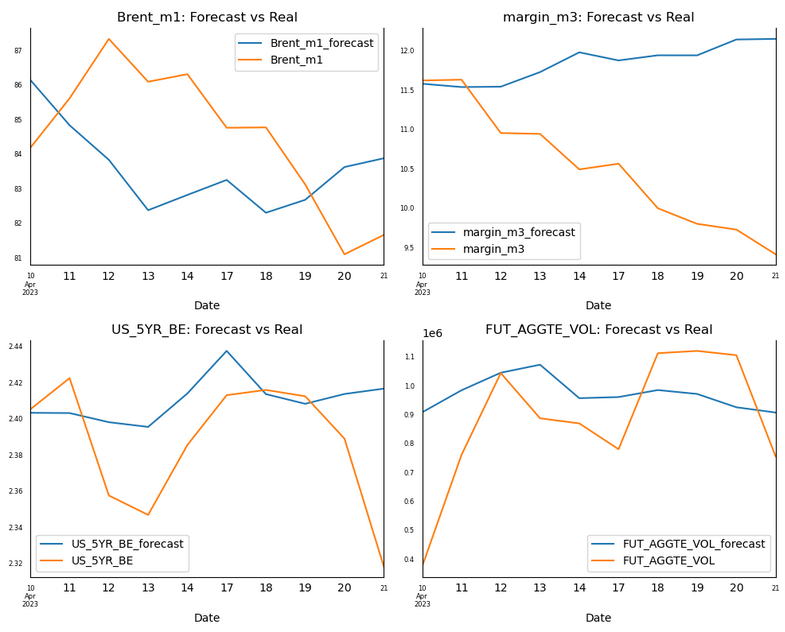
This program aims to develop a forecasting model based using Vector Autoregression. This was chosen based on the underlying assumption that the variables are dependent on each other. This multivariate timeseries model will relate the current observation of a variable with past observations of itself and the other variables in the data.

Firstly, the data will be pre-processed by removing data points with missing values and exploring the statistics. The data shows no missing values or other structural anomalies. A causation test is performed to check if the variables are causing each other. Using Granger’s Causality test, if the p-value of the Chi-squared test < 0.05, there is evidence that the variables are causing each other.

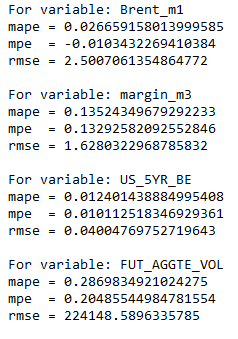
This shows the results, where it can be seen that there is sufficient evidence that there is causation in the given data. The volume shows little evidence of causation, which may suggest that it is not dependent on the other variables. Overall, this means that we can continue with the analysis, as regression analysis requires some causation.

Now, the data is split into train and test, where the last 10 values of the data is the testing set to be used as a benchmark against the forecasted values. Before fitting the model, the data needs to be checked for stationarity. If the data is nonstationary, the parameters of its distribution change over time. This can be removed by taking the difference of consecutive values in the data. This serves to remove the seasonality and other trends present in the data. An ADF test is performed to check the stationarity, where variables Brent\_m1, margin\_m3 and US\_5YR\_BE are nonstationary, while FUT\_AGGTE\_VOL is stationary. Differencing these yield all variables to be stationary, which is used as the input for the model.

A crucial step is to select the lag order, which is determined by minimising the AIC (Akaike information criterion). The VAR model is fitted with a range of lag values and the value with the lowest AIC is selected. For this data, it yields a value of 10. The model is now created and ready to forecast values. After the forecast is generated, it in the differenced state as a result of the stationary transformation earlier. After adding the trends back in, the results can be displayed.



The metrics for the forecast is shown below:



**Insights**

On inspection, the results are relatively positive with the trends displayed in the graphs above. It can be observed that the real data presents more volatility compared to the forecasted values, indicating the model is less suited for short term forecasts. For brent\_m1, the mape and mpe values indicate an error of around 2.6% and a small bias for underestimating the values. The rmse supports this, as the magnitude for the errors are relatively low. Similarly for margin\_m3 and US\_5YR\_BE have low values for the mape, mpe and rmse. This indicates an error of around 1% with a small bias for overestimate the values. The rmse for US\_5YR\_BE is particularly low, indicating a high precision in the forecasted values. For FUT\_AGGTE\_VOL, the high mape and mse indicate a large deviation between the forecasted and real values, with a high bias towards overestimation. Similarly, the rmse is high which suggests this model is unsuitable for predicting this particular variable.

From this, we can infer that while the brent\_m1, margin\_m3 and US\_5YE\_BE variables contain trends, FUT\_AGGTE\_VOL does not share the same property. Throughout this analysis, it has been clear that there is not significant evidence to suggest that the volume has any causation on the other variables. Applying these insights into the real world, this would make sense as while the volume can determine if the prices move significantly, it cannot reflect the direction of the movement. Similarly, it can be seen that the other variables do have some causation and therefore could be used in a model to predict future trends of each other.