WEST TEXAS INTERMEDIATE (WTI) CRUDE OIL ANALYSIS

Presenters:

Pratik Mulye

Rachit Jain

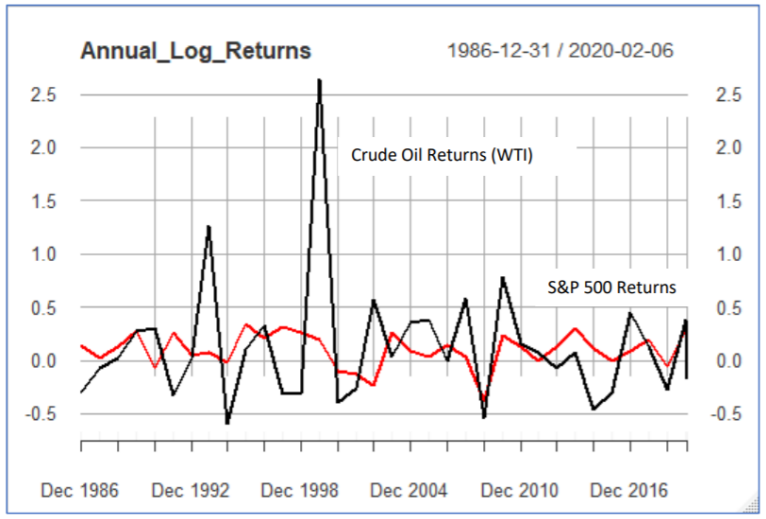
Bhakti Dalvi

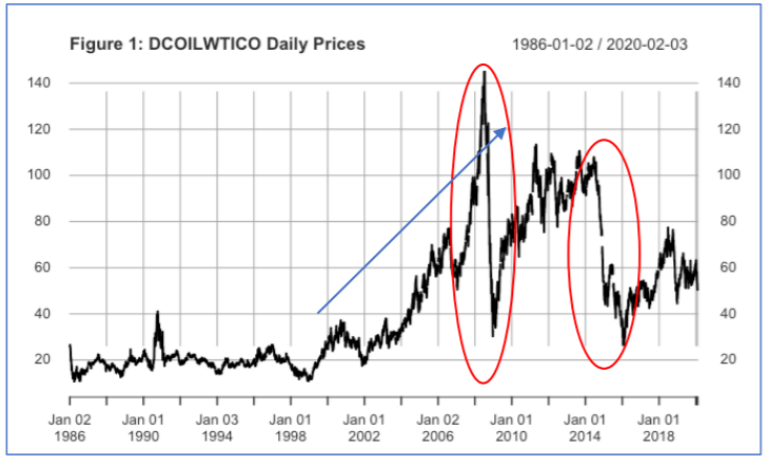
Nishika Jain

Background of the Asset & Market

Characteristics of WTI Market	WTI - Cushing is a crude stream produced in Texas and southern Oklahoma which serves as a reference or "marker" for pricing number of other crude stream.
Uses of Oil	manufacturing of clothes, for insulation and is of use in many Industries
Type of Crude Oil	WTI Crude Brent Crude
Factors that have an impact on crude oil	political crisis and wars, natural disasters, and the release of strategically invested oil reserves all influence the supply of crude oil and therefore it’s market price.

Background





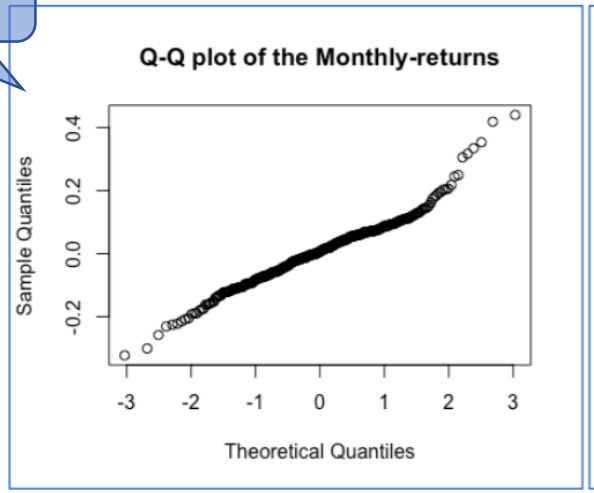
WHY WTI CRUDE OIL?

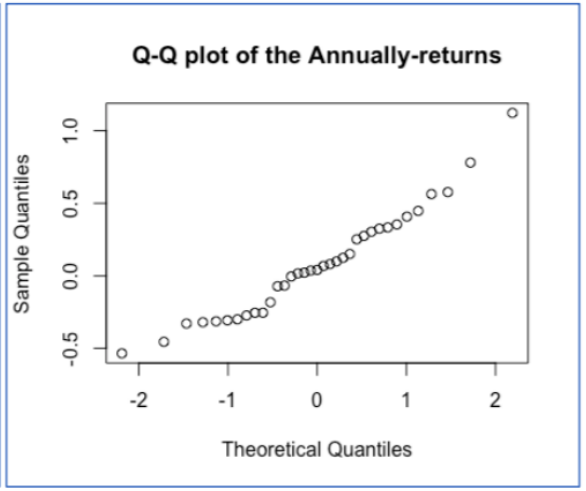
We select WTI Crude oil for our practice and econometric analysis. We have taken the prices from 1986 to present for our analysis. In this project (as a whole), we aim to understand how crude oil prices have been varying over the years (since 1986 till 2020), and attempt to forecast the next few months ahead prices using relevant model fitting and functions on R studio.

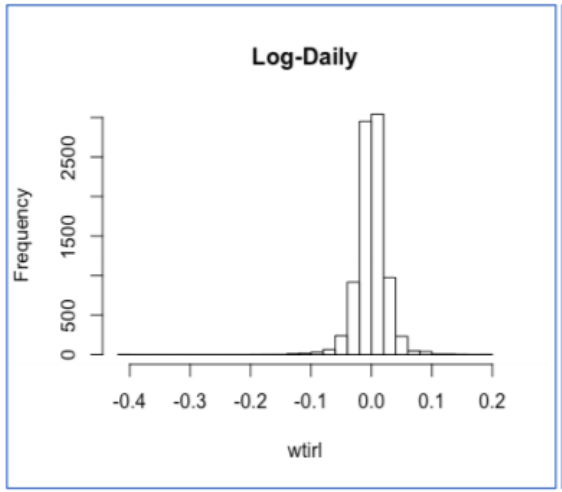
Data, Variables & Stylized Facts

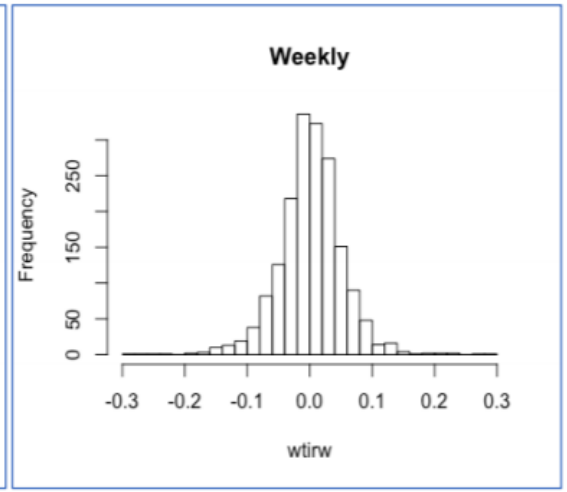
EXPLORATORY DATA ANALYSIS

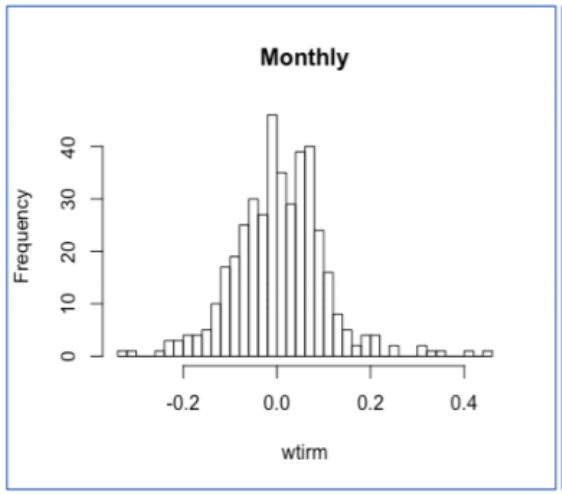
Our Selected Model

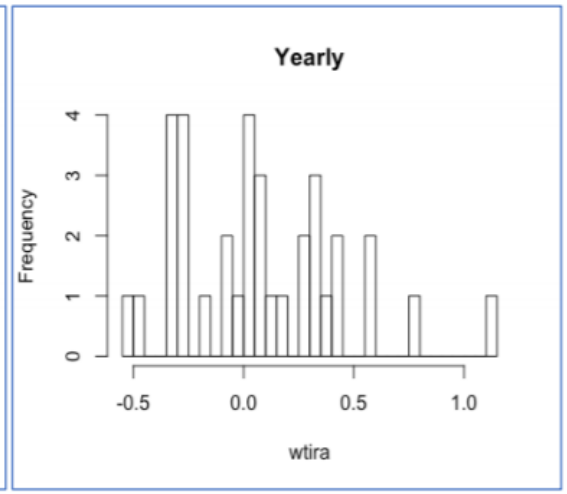


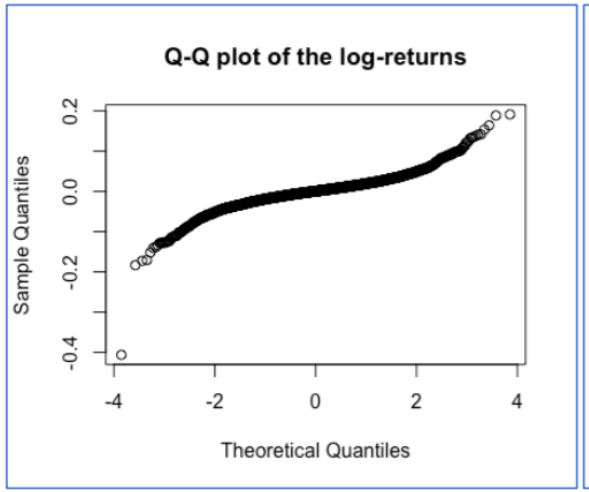


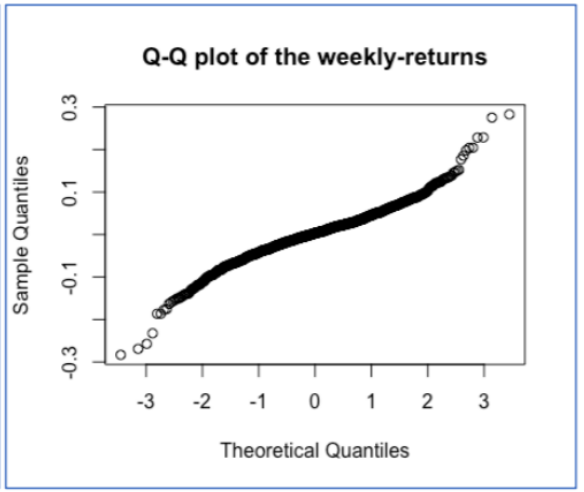






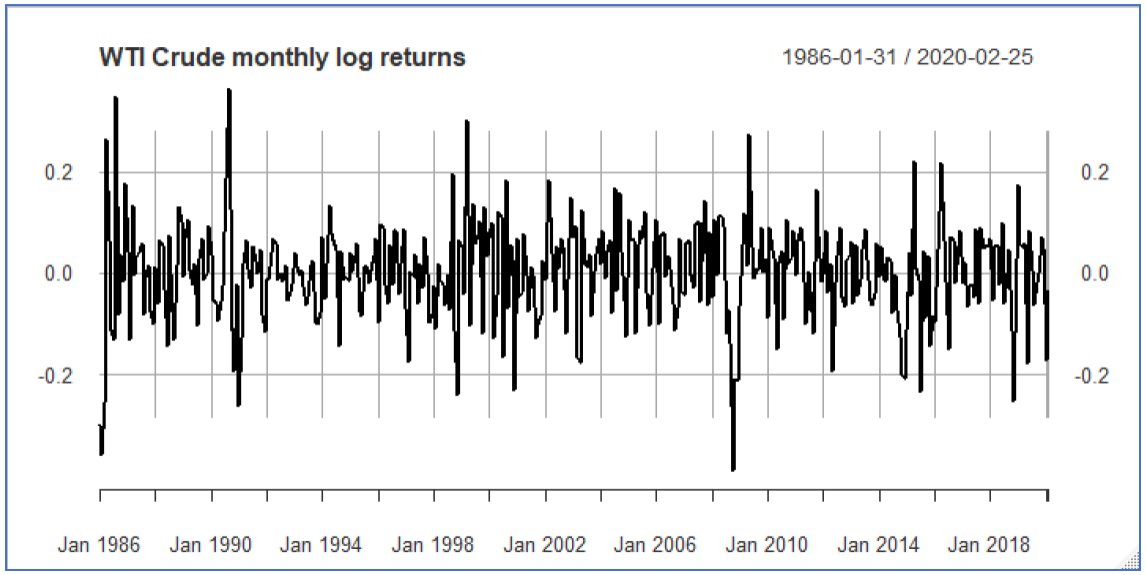


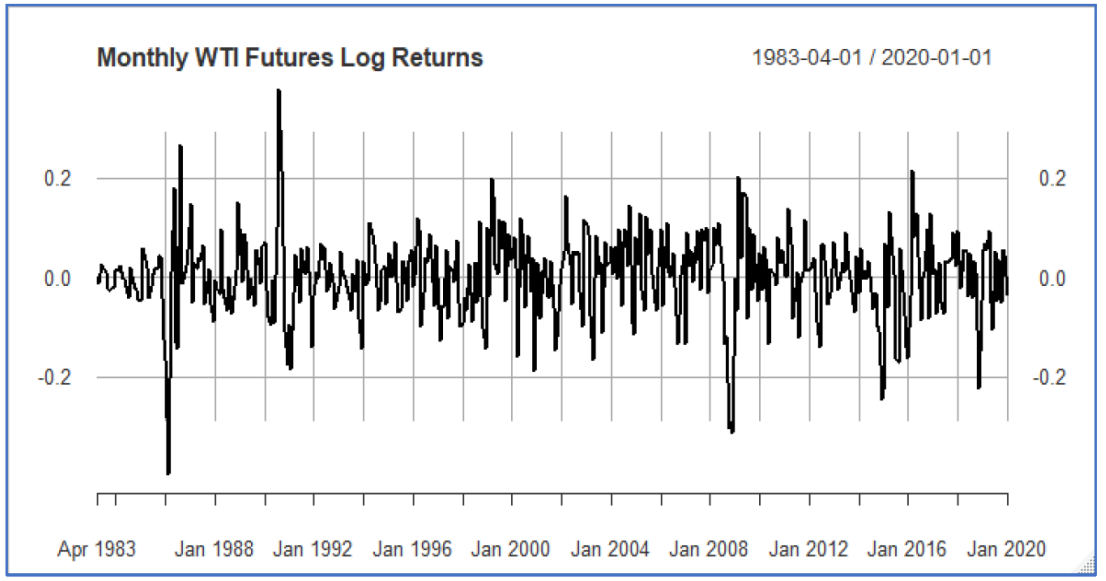




EXPLORATORY DATA ANALYSIS

Monthly Log Returns





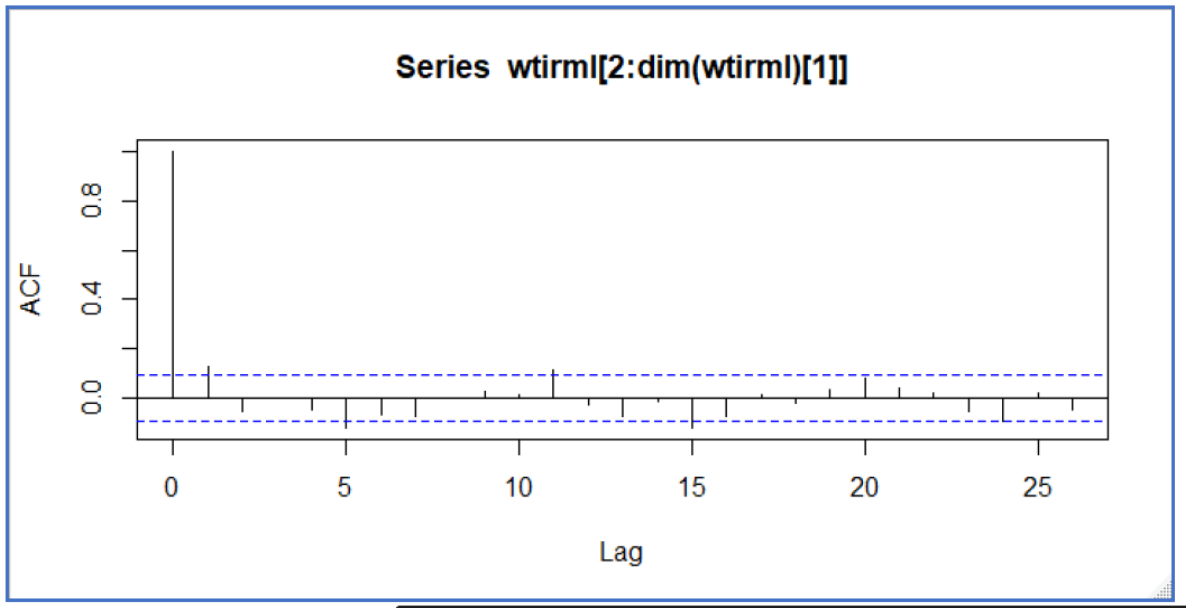
Weak Stationarity

Possibility of Bi-Variate analysis

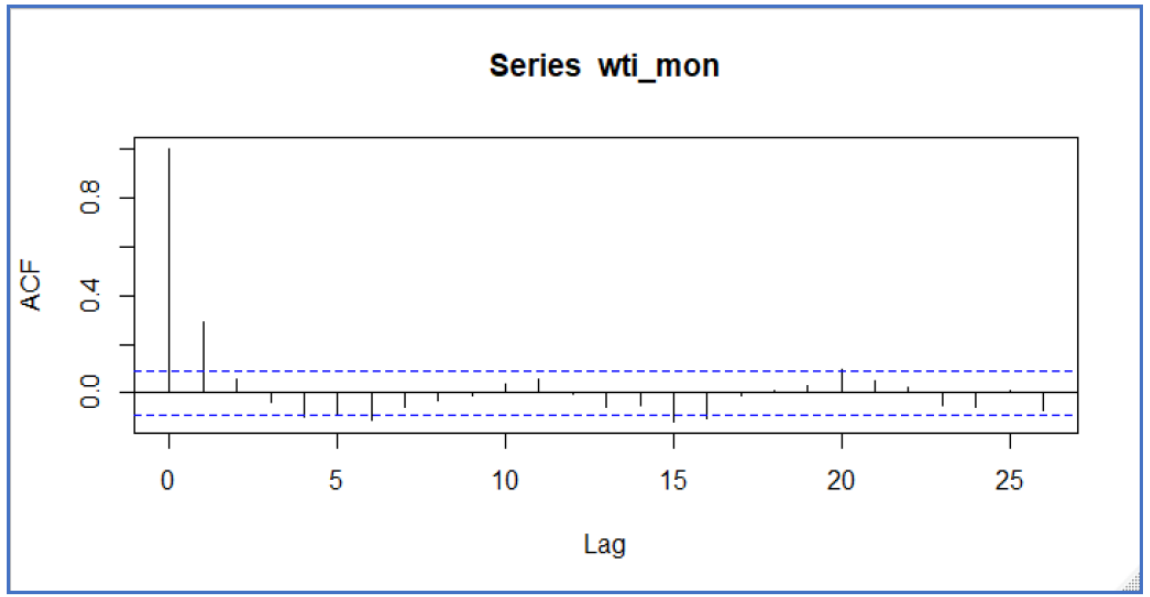
Dependency

EXPLORATORY DATA ANALYSIS

Monthly Spot Return



Monthly Futures Returns



We notice statistically significant correlation at Lag-1

A quick overview of econometric results

Basic Statistics:
Fat tails and skewed to left

Serial Correlation, hence, market inefficiency

No Unit-Root Problem

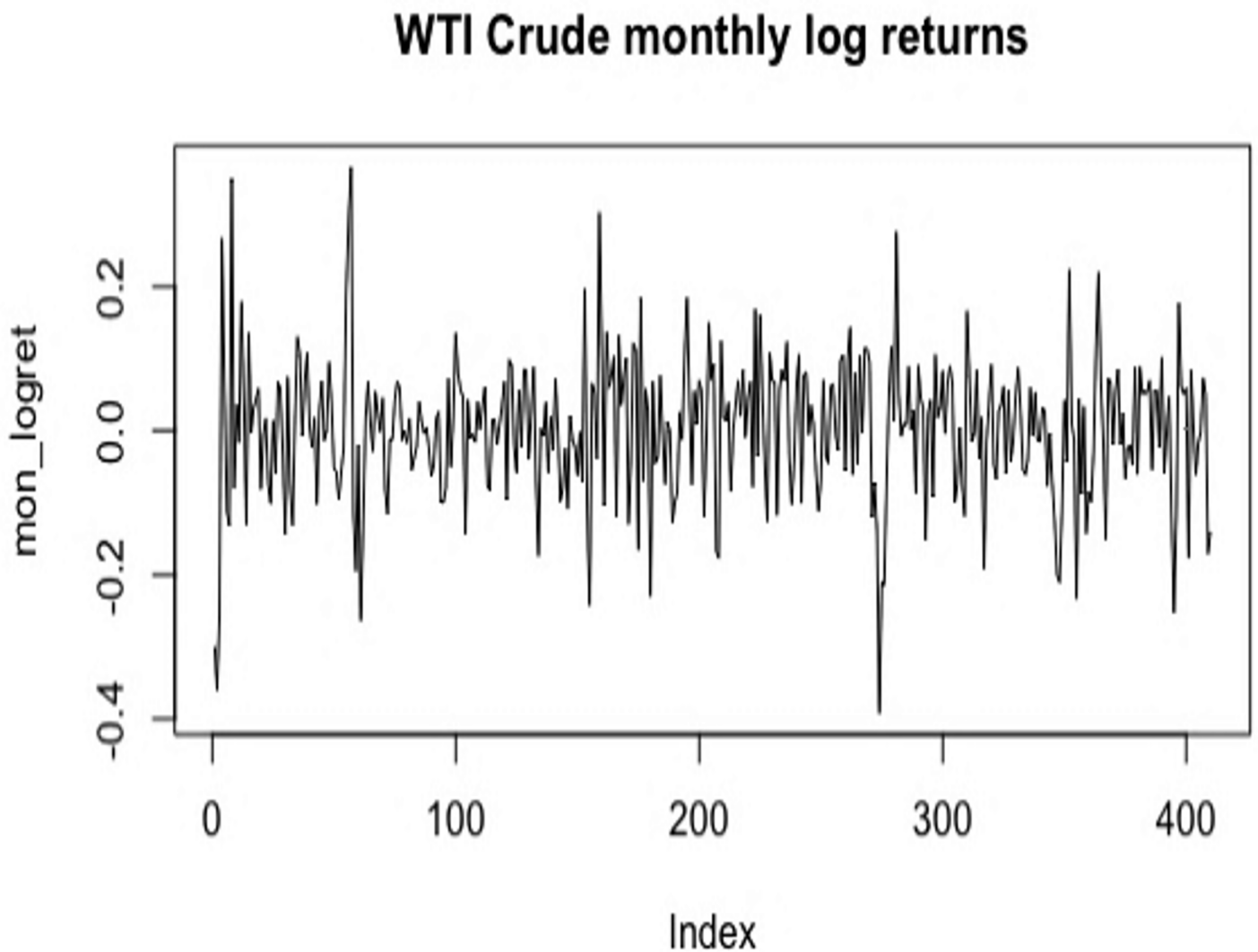
Seasonality existent

Heteroskedastic

Residuals can be standardized after GARCH fit

Forecasted returns for multi-periods

Value at Risk measurement

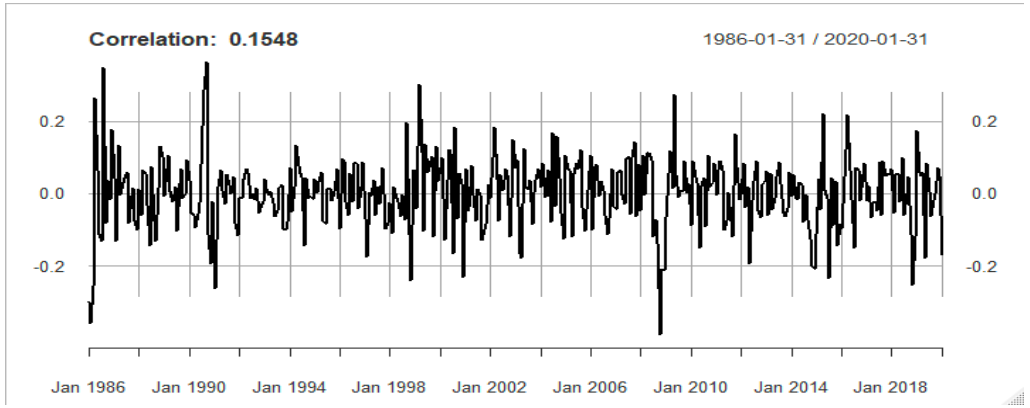


Uni-variate Modelling

Fitting an AR Model

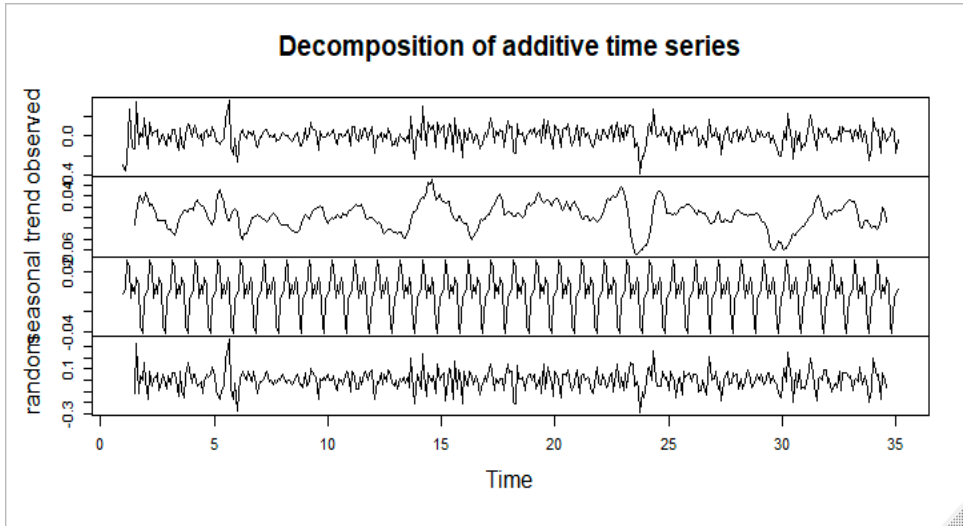
We found statistically significant serial correlation in the monthly log returns & hence the first trial was to fit an AR model

Using ACF & PACF for order determination, the appropriate order was at 1-lag



Auto-correlations were not very large between 1986 and 2018

Found seasonality component in the monthly log returns series

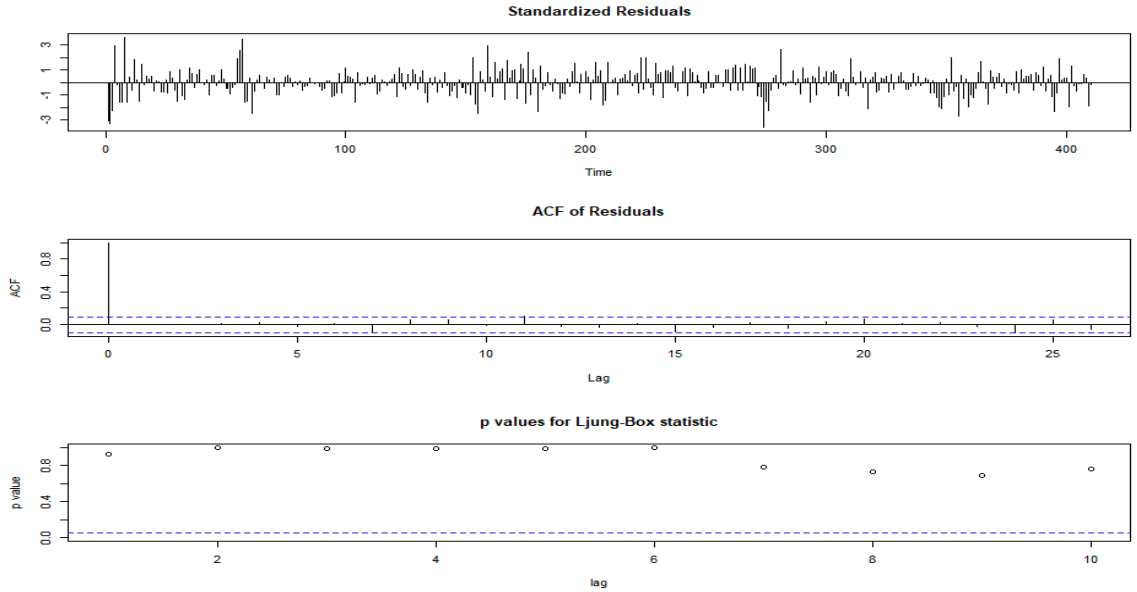


Shifted to ARIMA Model

Fitting an ARIMA Model

We used the auto ARIMA function to implement the model. Order (4,0,3) was found as optimal

```
Coefficients:
          ar1      ar2     ar3      ar4      ma1     ma2      ma3
      0.8747  -0.9548  1.0080  -0.2173  -0.7206  0.7913  -0.8879
s.e.  0.0840   0.0912  0.1225   0.0528   0.0718  0.0765   0.1230
```



The coefficients for all the factors were statistically significant and residuals that are normally distributed

Predicted Point	Predicted Value
At T+1	-0.008995914
At T+2	0.019516389
At T+11	0.003341314

Check for ARCH effect and heteroskedasticity in the series over time

Shifted to GARCH Model

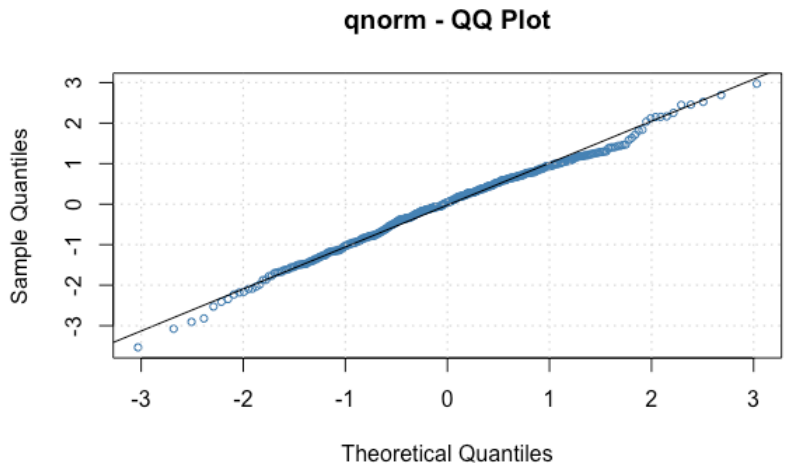
Fitting a GARCH Model

Implemented Mean Equation. ARCH effects test for a_t^2 for 10 lags

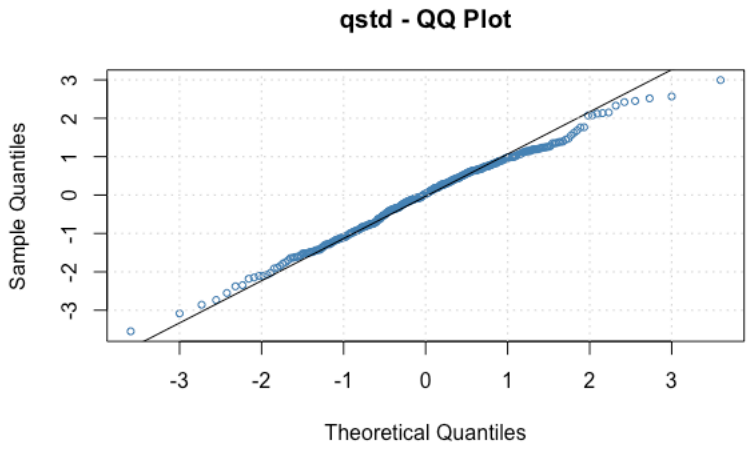
```
        Box-Ljung test

data:  at^2
X-squared = 94.629, df = 10, p-value = 6.661e-16
```

Rejected the Null-Hypothesis which shows weak ARCH effects. Still go for a GARCH model to check for normality of residuals



Much better fit with almost a standardized error series



Note slightly heavy tail. We fit a GARCH model with student t-distributions

Testing other GARCH Models

Fitting other GARCH Models

Cannot implement an IGARCH model as the coefficient is less than 1

Tested a GARCH-M model for relationship of return with volatility

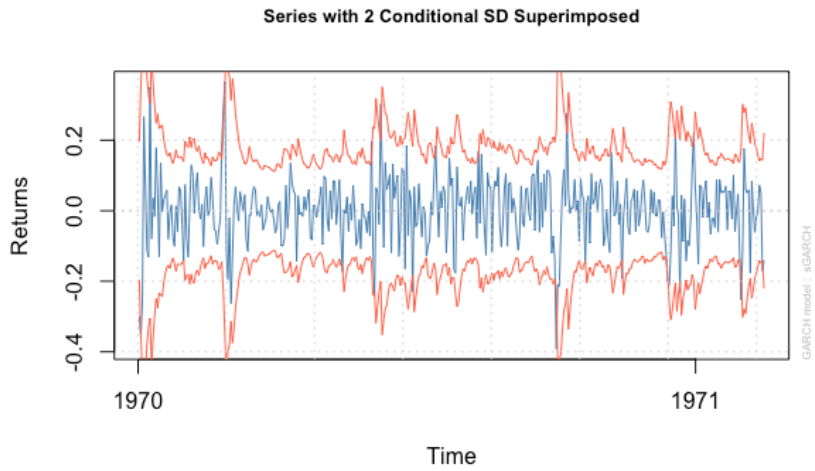
Coefficient(s):

	Estimate	Std. Error	t value	Pr(>|t|)	
mu	0.010245027	0.008527244	1.20145	0.22957804	
gamma	-0.901070338	1.044980101	-0.86228	0.38853088	
omega	0.001041168	0.000397416	2.61984	0.00879704	**
alpha	0.267902958	0.070284525	3.81169	0.00013802	***
beta	0.636724783	0.062655485	10.16232	< 2.22e-16	***

C parameter is negative implying negative risk premium

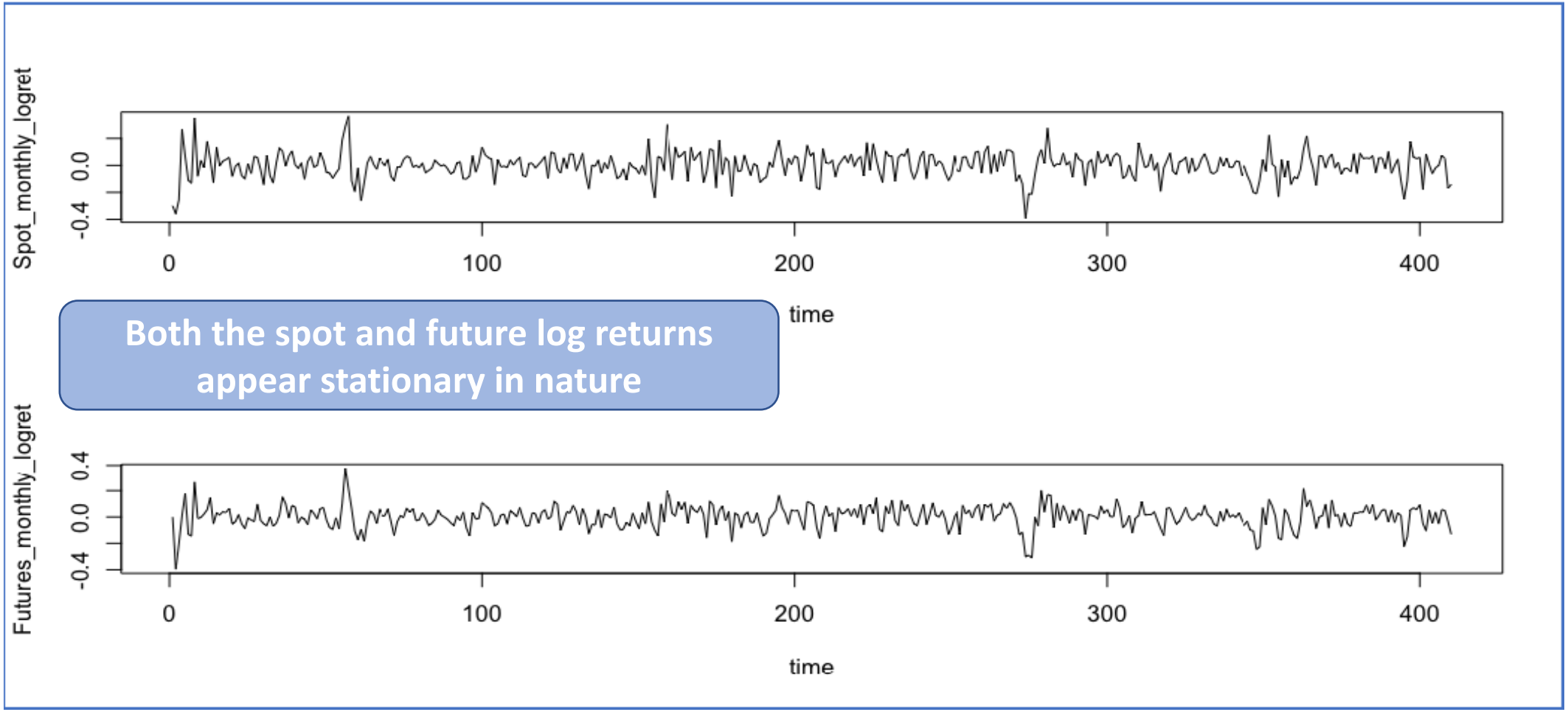
Tested E-GARCH and T-GARCH models for forecasting

Model was tested for telda a_t and found the residuals to be superimposed with 2 SD - normal distribution



A fitted GARCH Model

Multi-variate Modelling



Both the spot and future log returns appear stationary in nature

Stationarity Confirmed by the results of
ADF Test
(Augmented Dickey Fuller Test)

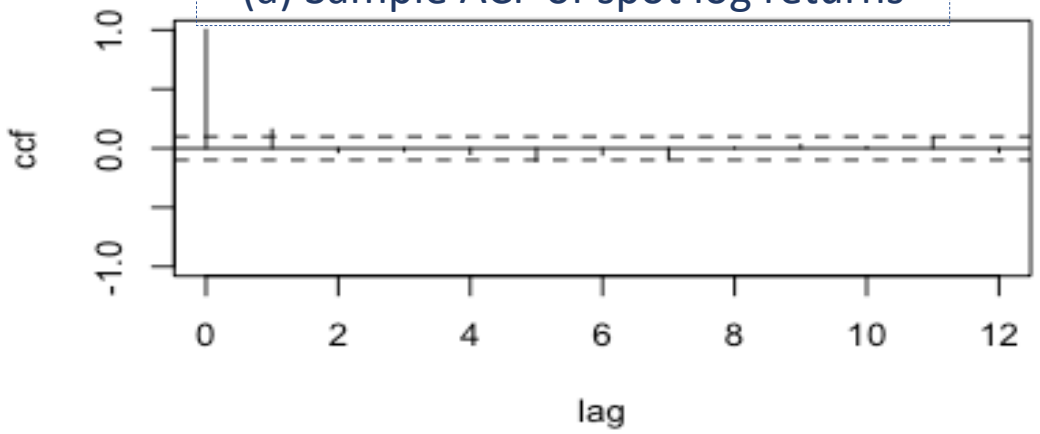
Augmented Dickey-Fuller Test

data: mon_logret
Dickey-Fuller = -8.1049, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

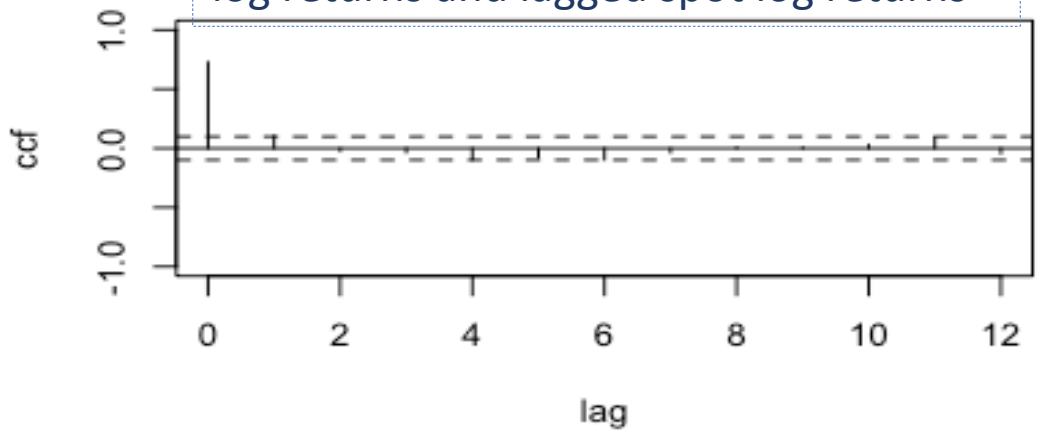
Augmented Dickey-Fuller Test

data: futures_logret
Dickey-Fuller = -7.9228, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

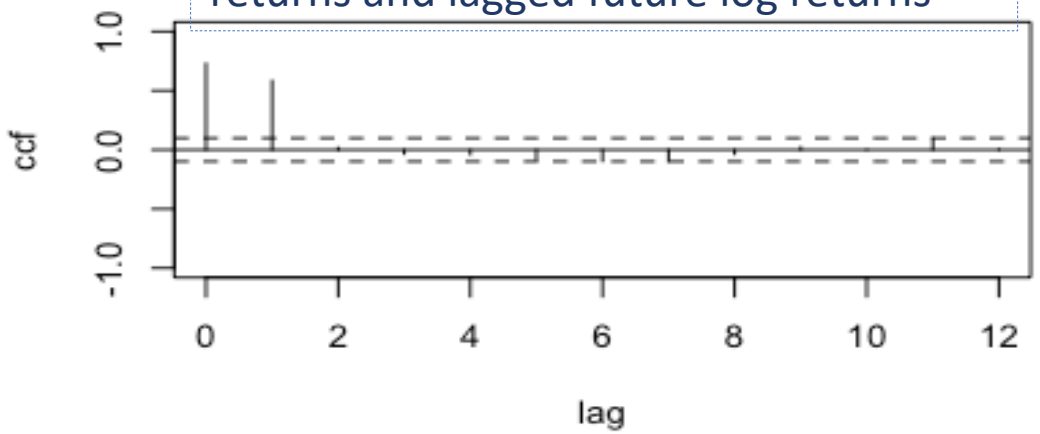
(a) Sample ACF of spot log returns



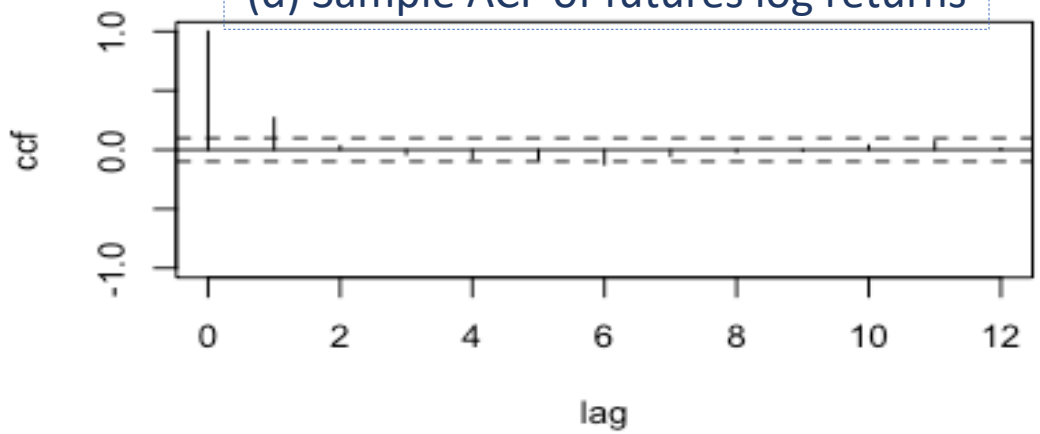
(b) Cross correlations between future log returns and lagged spot log returns

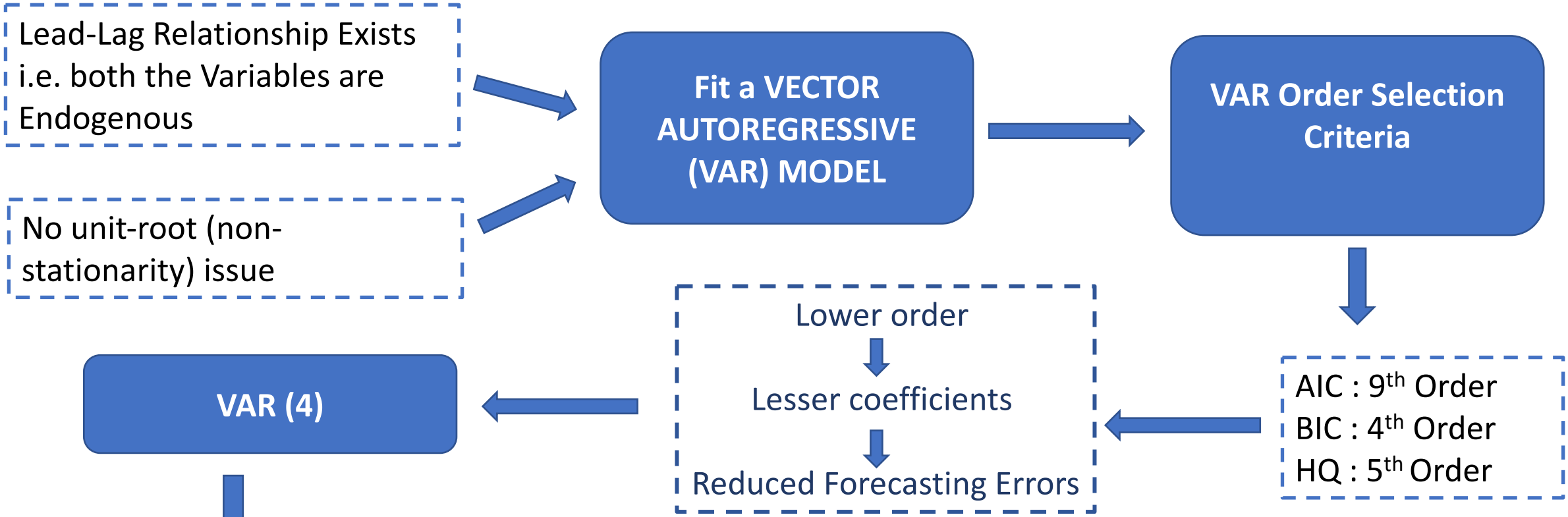


(c) Cross-correlations between spot log returns and lagged future log returns



(d) Sample ACF of futures log returns





	Spot Forecasts	Futures Forecasts	Prediction
T+0	-0.140256	-0.131430	Actual
T+1	0.004429	-0.10298	Forecasts
T+2	-0.007425	0.0137	Forecasts
T+3	-0.014473	-0.00576	Forecasts
T+4	0.022897	0.01232	Forecasts

Robustness of Results & Challenges in analysis

Our Econometric analysis performs well under regular market conditions but does not capture one-off movements. It captures trend, momentum and mean reversion but cannot address “Black Swan” event.

Dynamic sensitivities

Changes in delta, alpha, Mu etc.

Ambiguous Results between tools and results

Through:

- Confidence Intervals
- Statistical tests

Predicting extreme movements

Comparison with advanced Machine Learning techniques

Unsupervised Learning

Lessons, Conclusion & Future steps

Final Fitted Model

$$a_t = r_t - 0.00317$$
$$\sigma_t^2 = 0.0010 + 0.2679a_{t-1}^2 + 0.6367\sigma_{t-1}^2$$

The Econometric analysis helped to structure a base model for future enhancement of price & return predictability, also identifying risks in the market