

# Methodology

E B Benson

July 2024

## 1 Exploration and Pre-processing

The dataset was imported from a CSV file and initially explored using functions like `head()` and `summary()`. To enhance usability, column names were renamed. The date column was formatted correctly as a `Date` type, while the rate column was converted to numeric. Missing values were effectively managed through linear interpolation using the `na.approx()` function. A comprehensive plot was created to visualize the overall trend of the data.

## 2 Modeling and Diagnostic Techniques

To understand the data's autocorrelation structure, ACF and PACF plots were analyzed. A linear regression model was established as a baseline. An AR(1) model was subsequently tested, and an ARIMA(1,1,0) model was implemented to handle non-stationarity. Residuals were analyzed with the `checkresiduals()` function, and the Ljung-Box test checked for any remaining autocorrelation. To capture nonlinear dynamics, both TAR (Threshold Autoregressive) and STAR (Smooth Transition Autoregressive) models were applied. The Akaike Information Criterion (AIC) was used to compare model performance.

## 3 R Packages and Functions Used

For plotting, the `ggplot2` package was employed. The `zoo` package facilitated linear interpolation through the `na.approx()` function. The `forecast` package assisted with residual analysis via the `checkresiduals()` function. The `tsDyn` package was pivotal in fitting TAR and STAR models, utilizing `setar()` and `star()` functions, respectively. AIC values were calculated with the `AIC()` function, and forecasts were generated using the `predict()` function.