Investigating alternative selection criteria for ARIMA models.

Supervisor: Andriy Pototosky
Author: Tabish Ali Rather

Swinburne University of Technology

February 5, 2024

Research topic

ARIMA models are important in forecasting, and are useful for a wide range of disciplines from economics to environmental science. Despite their utility, selecting the optimal ARIMA model remains a challenge, notably in balancing complexity with forecasting accuracy. This research project aims explore alternative criteria for ARIMA model selection, directly comparing these methods with the widely recognised Hyndman-Khandakar algorithm. By emphasizing model performance while focusing on simplicity, this study seeks to enhance the practicality and efficiency of ARIMA modeling.

Data collection

For our project, we generated a dataset using a custom R script called `generate_random_arima`. This script is crucial for producing a broad spectrum of synthetic yet statistically valid time series data. The key aspects of this process include:

- Randomly selecting ARIMA parameters p, d, and q within pre-set boundaries to ensure a diverse dataset.
- Iteratively ensuring that each generated data set is stationary. This means the statistical properties of the series remain consistent over time.

This methodological step is essential for exploring the performance and complexity of ARIMA models selected using different criteria, thereby enriching the analytical depth of our research.

Theoretical Foundation and Unique Features

- **Selection criteria:** The algorithm selects models with maximum sum of p-values from SW, LBQ, and t-tests while Hyndman-Khandakar algorithm relies on AIC.
- **Theoretical Justification:**
  - *Shapiro-Wilk (SW)* test confirms residual normality, essential for model reliability.
  - *Ljung-Box Q (LBQ)* test checks for no auto-correlation in residuals, indicating a well-fitted model.
  - *T-test* assesses parameter significance, preventing overfitting.

  This approach ensures robust and generalisable ARIMA models by evaluating key model fit aspects.

Implementation and Comparative Points

- Implementation steps:
  1. Select the ARIMA model with the maximum sum of p-values from SW test, LBQ test, and t-test.
  2. Analyze the model to identify and remove overfit parameters.
  3. Refit the model without the overfit parameters and evaluate performance improvements.
- Comparative Points:
  1. Unlike traditional methods that may not systematically account for overfitting, this approach proactively identifies and corrects for it, potentially leading to more robust and generalisable models.
  2. The integrated use of multiple statistical tests to select the model distinguishes this method from others that might rely on a single criterion, offering a more holistic assessment of model fit.

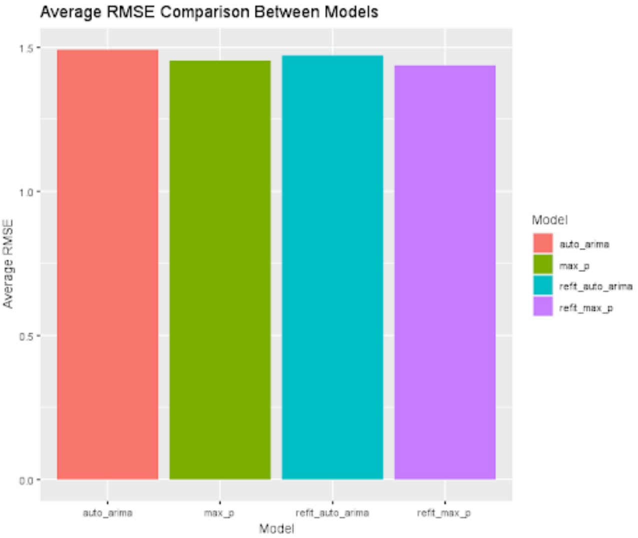
Key Comparison Criteria

- **Average RMSE:** Rolling window cross-validation on time series data to compute the Root Mean Square Error (RMSE), providing a direct measure of prediction accuracy. This method allows for a robust comparison of the predictive performance.
- **Insignificant Parameters Before and After Refit:** By evaluating the average number of statistically insignificant parameters before and after model refitting, we assess model complexity. This criterion reflects the optimisation by eliminating unnecessary parameters, thereby reducing complexity without compromising performance.

Context and rationale for comparison criteria

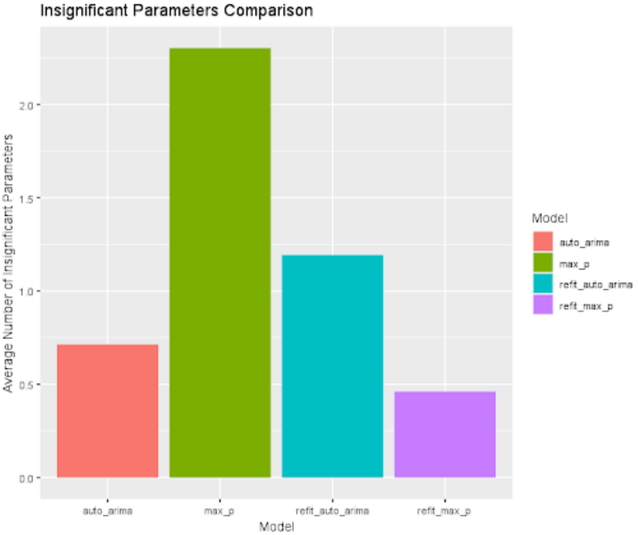
- **Contextualizing Metric Selection:**
  - The selection of Average RMSE and the assessment of Insignificant Parameters before and after refit are driven by the dual goals of forecasting accuracy and model simplicity.
  - These metrics provide a quantifiable means to evaluate the nuanced balance between model complexity and predictive performance, central to the effectiveness of ARIMA models in various applications.
- **Rationale Behind Metrics:**
  - **Average RMSE** offers a direct measure of how well each model forecasts new data, which is essential for real-world application of time series forecasting.
  - **Insignificant Parameters** evaluation reflects our commitment to model parsimony, crucial for computational efficiency and ease of interpretation without sacrificing accuracy.

Average RMSE Comparison



refit-max-p shows the best performance, i.e. the lowest RMSE value.

Comparison Average Number of Insignificant Parameters



Refit auto-max-p shows the best performance, lowest average number of insignificant parameters.