**DSO 424: Team Competition ( Due 1PM May 11, 2022)**

**Deliverables:**

1. **R code file(s) – submit on blackboard**
2. **10 min presentation on zoom – submit on blackboard and present 2-4pm on May 11 - present on zoom**

**Task:**

This assignment will ask you to build various models to forecast Private Housing Starts (PHS thereafter) according to the following scenarios:

* **Scenario 1:**  Monthly predictions: Training set <2007. Loop through all months in the testing set to predict next month’s value. Each time refit the model.
* **Scenario 2:**  12-month predictions: Training set <2007. Loop through all 12-month periods in the testing set to predict next 12 months value. Each time refit the model.
* **Scenario 3:**  Quarterly predictions: Training set <2007. Loop through all quarters in the testing set to predict next quarter’s value. Each time refit the model.
* **Scenario 4:**  Annual predictions: Training set <2007. Loop through all years in the testing set to predict next year’s value. Each time refit the model.

Example:

Naïve model for Scenario 1:

Iteration 1:

Build a model using all data up until December of 2006 (training set), predict for January of 2007(testing set, consist of 1 point)), evaluate RMSE and MAPE on both training and testing sets.

Iteration 2:

Build a model using all data up until January of 2007 (training set), predict for February of 2007(testing set, consist of 1 point)), evaluate RMSE and MAPE on both training and testing sets.

Iteration 3:

Build a model using all data up until February of 2007 (training set), predict for March of 2007(testing set, consist of 1 point)), evaluate RMSE and MAPE on both training and testing sets.

…….

Continue until you reach the most recent data point. Try as many models as you know. For each model calculate average MAPE on the testing set and present the results in a table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nr** | **Model** | **Average RMSE** | | **Average MAPE** | |
| **Training set** | **Testing set** | **Training set** | **Testing set** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| ….................. | | | | | |

Naïve model for Scenario 2:

Iteration 1:

Build a model using all data up until December of 2006 (training set), predict for January-December of 2007(testing set, consist of 12 points)), evaluate RMSE and MAPE on both training and testing sets.

Iteration 2:

Build a model using all data up until January of 2007 (training set), predict for February of 2007-January of 2008(testing set, consist of 12 points)), evaluate RMSE and MAPE on both training and testing sets.

Iteration 3:

Build a model using all data up until February of 2007 (training set), predict for March of 2007-February of 2008(testing set, consist of 1 point)), evaluate RMSE and MAPE on both training and testing sets.

…….

Continue until you reach the most recent data point. Try as many models as you know. For each model calculate average MAPE on the testing set and present the results in a table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nr** | **Model** | **Average RMSE** | | **Average MAPE** | |
| **Training set** | **Testing set** | **Training set** | **Testing set** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| ….................. | | | | | |

Naïve model for Scenario 3:

Convert monthly data to quarterly : Q1 data = January data +February data +March data, etc. and then apply the approach similar to Naïve model for Scenario 1 for quarterly data.

Naïve model for Scenario 4:

Convert monthly data to annual: yearly data = January data +February data +…+ December data, etc. and then apply the approach similar to Naïve model for Scenario 1 for annual data.

**Instructions:**

1. Import PHS data from FRED directly into R:

[**https://fred.stlouisfed.org/series/HOUSTNSA**](https://fred.stlouisfed.org/series/HOUSTNSA)

1. Apply a variety of forecasting techniques to predict housing starts for each Scenario described above. Present the results in a table format for each scenario (a total of 4 tables: one table for each scenario). Include all 4 tables in your presentation.

Text, letter

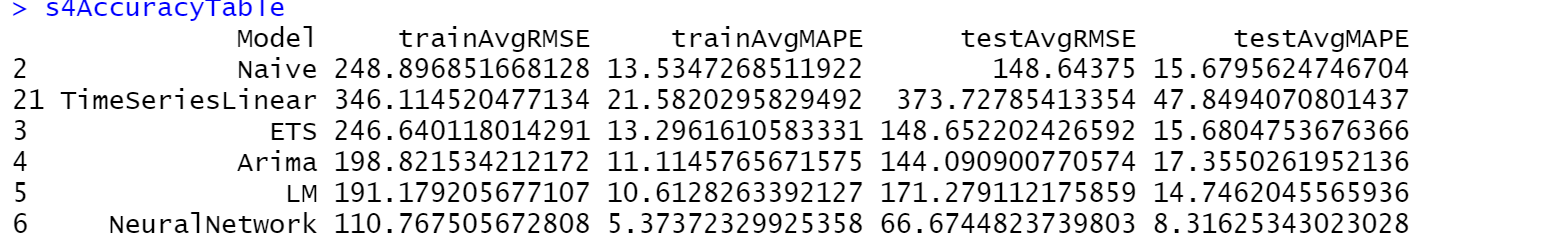
Description automatically generated

Text, letter

Description automatically generated

Text

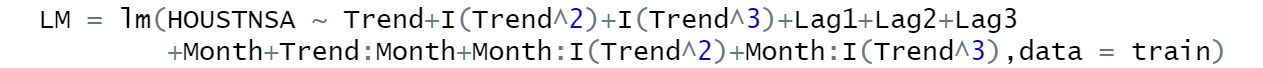
Description automatically generated



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Nr** | **Model** | **Average RMSE** | | **Average MAPE** | |
| **Training set** | **Testing set** | **Training set** | **Testing set** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| ….................. | | | | | |
|  | | | | | |

1. Create visualizations for each scenario. I let you decide on what visualizations to create. Focus visualizations during your presentation.

Scenario1: (Arima Model)



Chart

Description automatically generated(at 2006 dec, forecast 2007 Jan and later)

Chart

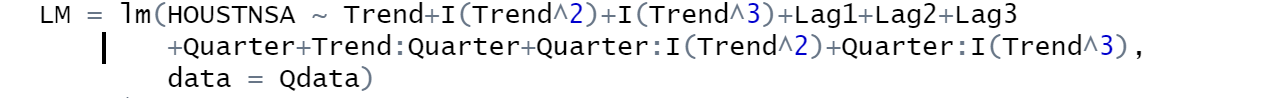
Description automatically generatedeach predication point

Scenario2: (LM model)

Chart

Description automatically generated

Scenario3:



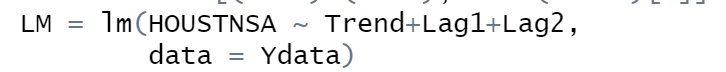
Chart

Description automatically generatedLM

Chart

Description automatically generatedLM model

Scenario 4: at time 2006, forecast 2007 and later on



Chart, line chart

Description automatically generatedLM

1. What model would you recommend an economic consulting start up to use for forecasting private housing starts?

Text

Description automatically generated

Text

Description automatically generated

LM

1. **Extra credit:** Research housing industry and identify leading indicators to predict PHSCF. Do they improve the accuracy of PHSCF forecasts?   
   Extra credit is based on an individual effort. There is a separate submission link to submit your R code file.

To improve the forecast, we can divide the housing market into smaller segments, for example, high-price band, middle-price band, and low-price band (or by geographical regions). With predictions from each finer categories weighted and added into the final housing trends, the forecasts can be more precise.

Last semester, my group and I conducted research on LA individual housing prices using economic model. We include housing features, such as the number of beds, area of houses, and ages built, and parameters measuring the community from education, medical service, safety factors, entertainment, economic level, and population density (zip code or city). So, the independent variables include number of baths and bedrooms in the house, lot size, area of the house in square feet, year built, total crime in the zip code, the estimated number of homeless (per mi sq), number of food stamp, graduate enrollment, undergraduate enrollment, park space acres per capita, economic hardship index, hospital dummy, good schools dummy (0 or 1), median income, house type, and population density (per mi sq). The regression can capture 82.6 per cent of the variations in individual housing prices.

Although economic model cannot be a useful way of prediction as the independent variables are unknown in the future, it can lead us to think more of making predictions for each finer category and sum it up for a better prediction.

A close-up of a document

Description automatically generated with medium confidence