Week 6 Assignments

# Monthly Carbon Dioxide Levels at Mauna Loa (50 Points)

# Description

Monthly mean carbon dioxide (in ppm, parts per million) measured at Mauna Loa Observatory, Hawaii. This is an update to CO2 in the datasets package.

# Format

Original format is: Time-Series [1:781] from 1958 to 2023: 316 317 318 317 316 ...

New format is: Data-frame, name: cardox.csv

# Details

The carbon dioxide data measured as the mole fraction in dry air, on Mauna Loa constitute the longest record of direct measurements of CO2 in the atmosphere. They were started by C. David Keeling of the Scripps Institution of Oceanography in March of 1958 at a facility of the National Oceanic and Atmospheric Administration. NOAA started its own CO2 measurements in May of 1974, and they have run in parallel with those made by Scripps since then. Data are reported as a dry mole fraction defined as the number of molecules of carbon dioxide divided by the number of molecules of dry air multiplied by one million (ppm).

Due to the eruption of the Mauna Loa Volcano, measurements from Mauna Loa Observatory were suspended as of Nov. 29, 2022. Observations starting in December 2022 are from a site at the Maunakea Observatories, approximately 21 miles north of the Mauna Loa Observatory.

# Source

<https://gml.noaa.gov/ccgg/trends/>

# Forecasting Process

1. Read the co2.df
2. Prepare proper timeseries dataset

A graph on a computer screen

Description automatically generated

1. What month in all years has the highest level of CO2?

A graph of carbon dioxide levels

Description automatically generatedA graph of the seasons

Description automatically generated with medium confidenceThe below seasonal plot shows timeseries for every year, For clear vision of plot to find which month in all years has the highest level of CO2 I have created seasonal plot for first five years to see clearly. **May Month** has the highest level of CO2 for all years.

1. Timeseries exploration
   1. Use the decompose () to Extract the residuals, trend, and seasonality values from timeseries data

A graph of different types of time

Description automatically generated with medium confidence

* 1. Rebuilt a dataset by summing these components.
  2. Plot the original time series and new data. Explain your visualization.
* The plot illustrates a close alignment between the original carbon dioxide data and the rebuilt dataset, suggesting successful decomposition and reconstruction processes. This alignment indicates that the trend, seasonal, and random components effectively capture the underlying patterns of the original time series.

A graph with green dots

Description automatically generated

1. Building Models
   1. Partition timeseries
   2. Training from 1956 March to 2019 March
   3. Validation 2019 March to 2023 March
   4. Build three type of regression models
      * Linear regression
      * Parabolic regression
      * Exponential regression
   5. Build triple exponential smoothing with ets() use an MAA setting
2. Forecast validation with these models and compare their performance measures.

* **From the accuracy measures, it appears that the Triple exponential smoothing model (ETS) performs the best among the evaluated forecasting models.**

A screenshot of a computer program

Description automatically generated

1. Prepare required plots. DO not generate irrelevant plots.

A graph showing a number of different types of models

Description automatically generated with medium confidence

1. Write summary of your understanding of this case in couple of lines

This case study explores monthly CO2 data from Mauna Loa, identifying seasonal patterns through visualization. Next, we decomposed the data into trend, seasonal, and random components and rebuilt the original dataset. Then, we partitioned the data for model training and validation, building regression models (linear, parabolic, exponential) and employing triple exponential smoothing for forecasting. Finally, we evaluated model performance using accuracy measures and visually compared forecasted values against the original data and validation sets.

1. Manually forecast for December 2023 and January 2024 with your best regression model.

* **From three regression models, Parabolic regression model performed best. I will use this model on the entire dataset to forecast December 2023 and January 2024.**

A close-up of numbers

Description automatically generated

* Dataset with quadratic trend and additive seasonality

Where are trend and

are seasons.

are coefficients of seasons.

Here the time(t) is calculated from the original dataset, which has total no of rows 781 and last month is march. So, for March 2023 the time is 781, Then for December 2023 the time is 790.

* **For December 2023, t =790**

= 314.595 + 0.06368\*(790) + 0.00009115\*( + 0.6402(0) + 0 +…. + 0 + (0.94654987) \*1

= 314.595 + 50.3072 + 56.886715 – 0. 94654987

= 420.842365

* **For January 2024, t = 791**= 314.595 + 0.06368\*(791) + 0.00009115\*( + 0.6402(0) + 0 +…. + 0

= 314.595 + 50.37088 + 57.0308232

= 421.996703

* **Forecast using the model in R and Manual is approximately same.**

A screenshot of a computer

Description automatically generated

# Hawaiian Occupancy rates (25 Points)

# Description

Quarterly Hawaiian hotel occupancy rate (percent of rooms occupied) from 1982-I to 2015-IV

# Formatyes

Original format is: Time-Series [1:136] from 1982 to 2015: 79 65.9 70.9 66.7 ...

New format is: Data-frame, name: hor.csv

# Source

https://dbedt.hawaii.gov/economic/qser/tourism/

# Forecasting Process

1. Read the hor.df
2. Prepare proper timeseries dataset
3. Timeseries exploration
   1. Prepare the decompose plot of timeseries. Interpret timeseries components

**Time series components**

**Trend:** There is no trend in this time series data.

**Seasonality:** There is Quarterly seasonalityA graph of different types of waves

Description automatically generated with medium confidence

* 1. Plot autocorrelation and partial autocorrelation for 12, 36, and all periods.

A group of graphs showing the number of the same period

Description automatically generated with medium confidence

1. Building models
   1. Partition timeseries
   2. Training from 1982 January to 2009 December
   3. Validation 2010 January to 2015 December
   4. Build three type of regression models
      * Build an ets() model with proper model setting.
      * Build an autoregressive model with residuals
2. Forecast validation with ets() models and AR().
3. Forecast for three years (2016-2019) into the future using a model built with ets()

A screenshot of a computer screen

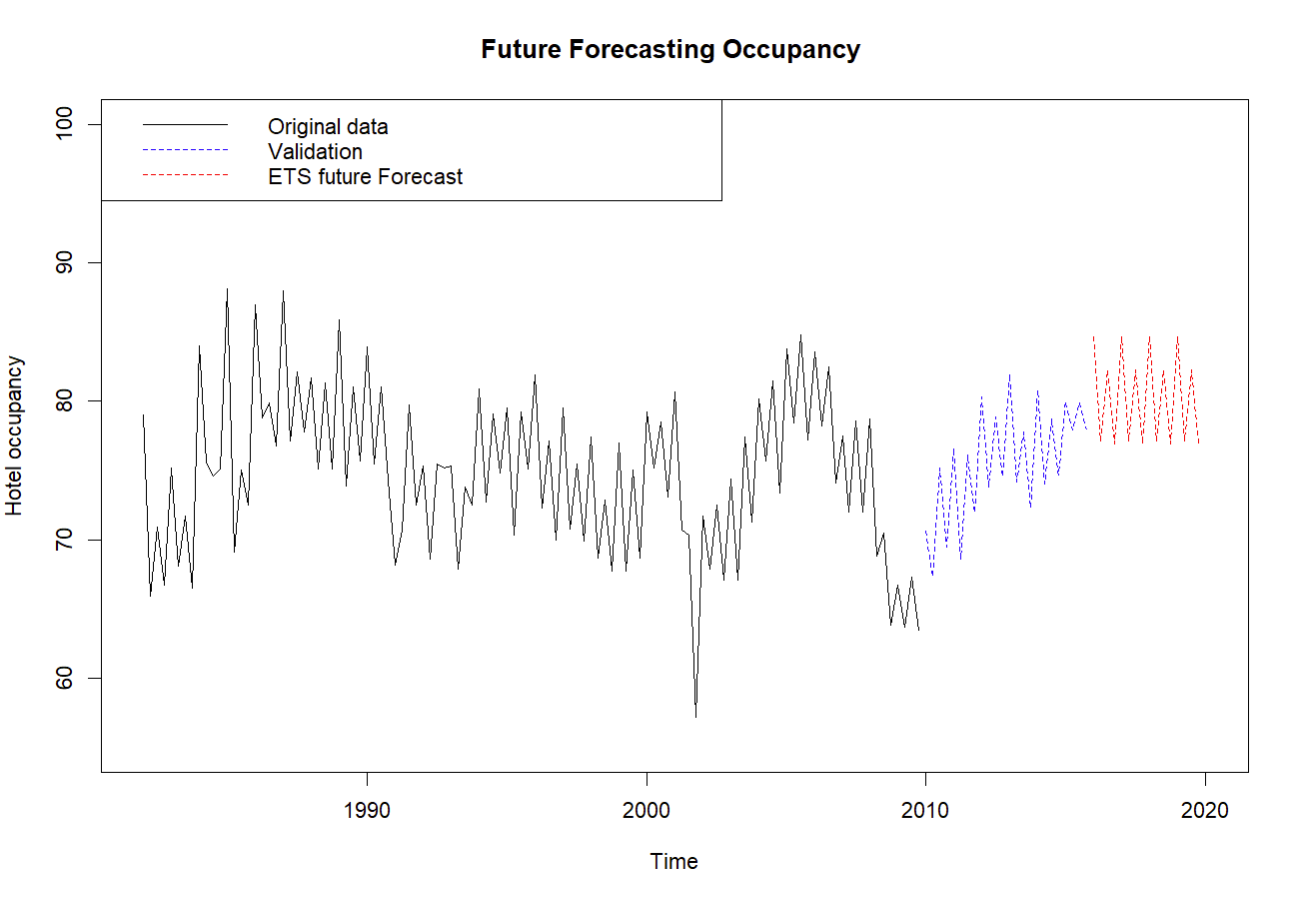
Description automatically generated

1. Prepare a plot of
   1. Training timeseries
   2. Validation timeseries
   3. The ets model forecast.
   4. The AR forecast
   5. Sum of ets and AR forecast values

A graph showing a number of numbers

Description automatically generated with medium confidence

1. Prepare a plot of the future forecast.



1. Write summary of your understanding of this case in couple of lines

The case involves exploring hotel occupancy timeseries data, building and validating forecasting models, and generating future forecasts. The ETS model appears to be effective for forecasting hotel occupancy, as demonstrated by its performance on the validation data and future forecasts.

**Question 3:**

1. (25 Points) Do all parts of problem 3 on page 172.
   1. Run a regression model with log (Sales) as the output variable and with a linear trend and monthly predictors. Use this model to forecast the sales in January 2001, January 2002, and February 2002. Think carefully which data to use for model fitting in each case.

**Forecasted values are approximately the same in both R and Manual calculations.**

* 1. Using the training period, create an ACF plot until lag-15 for the forecast errors. Now fit an AR model with lag-2 [AR (2)] to the forecast errors.
* Examining the ACF plot and the coefficients of the AR(2) model (and their statistical

significance), what can we learn about the regression model forecasts?

**Arima Model coefficients**

A graph with lines and numbers

Description automatically generated

A number and text on a white background

Description automatically generated

**ACF Plot Interpretation:**

The ACF plot reveals significant autocorrelation at lag -1 and lag -2, as indicated by autocorrelation coefficients outside the bounds (blue dotted lines). Additionally, for the first 10 lags, there is only positive autocorrelation, but from lag 11 onwards, there is only negative autocorrelation.

* Compute forecasts for January 2001, January 2002, and February 2002, using the regression and AR (2) index {AR(2)} models.

Again, think which data can be used for fitting the regression and AR(2) models in each case

**Add to the part b the following:**

Manually, calculate the sales for January 2001, January, February, and March of 2002. No R code should be involved. Show all the steps of your calculations.

**Regression Model:**

* A screenshot of a computer

  Description automatically generated**For January 2001, t =73**

Dataset with linear trend and additive seasonality (assuming our season is M and we have N seasons)

= 7.646363 + 0.021120 \* 73 + 0 + 0 … + 0

= 9.188123

* **Forecast in R**



**A number and number in a row

Description automatically generated with medium confidence**

* **Consider the entire data to fit the regression model and forecast.**
* **For January 2002, t = 85**

= 7.6058604 + 0.0223930 \* 85 + 0 +… + 0

= 9.5092654

* **For February 2002, t =86, M2=1**

= 7.6058604 + 0.0223930 \* 86 + 0.2510437\*1 +0+… + 0

= 7.6058604 + 1.925798 + 0.2510437

= 9.7827021

* **For March 2002, t =87, M3=1 Forecast in R**

A black and white text

Description automatically generated= 7.6058604 + 0.0223930 \* 87 + 0.2510437\*0 + 0.6952066 \* 1+… + 0

= 7.6058604 + 1.925798 + 0.6952066

= 10.226865

* **ARIMA Model:**

**Forecast using the Arima model in R.**

**January 2001:**

****

**Year 2002:**

A number and date on a white background

Description automatically generated