Assignment-4

Data Analysis and Interpretation

Nischal 23b1024

Pushpendra Uikey 23B1023

Nithin 23B0993

Professor: **Sunita Sarawagi**

October 28, 2024

1 Parking Lot Problem

- **Part (a):**
  - Mean Absolute Percentage Error (MAPE): 5.04%
  - Mean Absolute Scaled Error (MASE): 0.80
- **Part (b):**
  - Mean Absolute Percentage Error (MAPE): 1.85%
  - Mean Absolute Scaled Error (MASE): 0.27

2 Analysis of Forecasting Metrics on a Real Dataset

2.1 Limitations of MAPE as a Metric

The Mean Absolute Percentage Error (MAPE) is often chosen as a metric for forecast accuracy. It measures the percentage discrepancy between actual and forecasted values:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{A_i - F_i}{A_i} \right| \times 100 \tag{1}$$

where:

- A_i denotes the actual observed values,
- F_i represents the forecasted values, and
- n is the count of observations.

2.1.1 Challenges with MAPE

- **Instability with Low Values:** For data points with very low actual values, MAPE can produce extremely high percentages or even become undefined, distorting the error measurement and making it less reliable.
- **Skewed Emphasis on Small Values:** Since MAPE calculates error inversely proportional to actual values, it can unfairly magnify errors for lower values, causing unbalanced error representation across different demand levels.
- **Minimal Impact on Peak Periods:** High-demand periods, often critical for resource allocation, might not be adequately emphasized by MAPE since it does not inherently prioritize errors based on demand magnitude.

2.2 Alternative Metric: RMSE

To address MAPE's limitations, the Root Mean Squared Error (RMSE) offers an alternative by emphasizing larger errors. RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - F_i)^2} \tag{2}$$

2.2.1 Advantages of RMSE

- **Focus on Significant Errors:** RMSE increases the weight on larger errors, making it more effective at reflecting significant deviations, which is beneficial when forecasting for resource-sensitive applications.
- **Suitability for Planning Scenarios:** By providing a comprehensive error measure, RMSE is useful in settings where larger forecast deviations could lead to substantial planning or resource allocation issues.

2.2.2 Example Calculation

Let's consider the following data points to compare MAPE and RMSE more concretely:

Month	Actual Count (A_i)	Forecasted Count (F_i)
Month 1	100	80
Month 2	5	10

Table 1: Sample Actual vs Forecasted Values

For MAPE:

$$\text{MAPE} = \frac{1}{2}\left(\frac{|100-80|}{100} \times 100 + \frac{|5-10|}{5} \times 100\right) = 60\%$$

For RMSE:

$$\text{RMSE} = \sqrt{\frac{(100-80)^2 + (5-10)^2}{2}} = \sqrt{\frac{400+25}{2}} = \sqrt{212.5} \approx 14.58$$

In this case, RMSE provides a less exaggerated view of the errors, better aligning with practical needs.

2.3 Assessing Pre-COVID vs. Post-COVID Differences

To understand the potential shift in trends across different time periods, particularly around the COVID-19 period, we examine the first differenced series, ΔY, which is assumed to be weakly stationary.

2.3.1 Method Selection: Two-Sample t-Test

A two-sample t-test is appropriate to determine whether there's a statistically significant difference in mean values of ΔY across two periods: pre-COVID (before December 2019) and post-COVID (after January 2022).

Hypothesis Formulation:

- **Null Hypothesis** (H_0): The average of ΔY remains consistent across both periods, implying no significant difference.
- **Alternative Hypothesis** (H_a): The averages of ΔY differ between the two periods, indicating a significant shift.

Example of Calculation: Assuming hypothetical values for pre- and post-COVID periods, the following table summarizes the differenced data:

Period	First Differences (ΔY)
Pre-COVID	5, 7, 8, 6, 5
Post-COVID	2, 3, 1, 4, 3

Table 2: Sample First Differences by Period

Calculated mean values:

- Pre-COVID Mean: $\frac{5+7+8+6+5}{5} = 6.2$
- Post-COVID Mean: $\frac{2+3+1+4+3}{5} = 2.6$

Using statistical software or t-table values, the test determines if there is a significant p-value indicating a shift in averages between these periods.