Univariate Time Series Data and Model Card

Generated by Cardtale - Automated Model and Data Card Generator

This report provides an automated, comprehensive analysis of univariate time series data. Generated by Cardtale, it explores basic aspects and potential challenges in your data to support informed decision-making and modeling choices.

Generated: 2024-11-01 09:42

Series Name: M101

Table of Contents

Other aspects were explored but omitted from the final report:

Variance

Hypothesis testing suggests that the time series has constant variance (homoskedasticity). In preliminary tests, common transformations for variance stabilization did not improve forecasting accuracy

Change Detection

No change point was found according to offline change detection methods

Data Overview

This section examines the core characteristics and statistical properties of the time series. Understanding these attributes is important for assessing data quality and gaining a preliminary context. We explore the temporal structure, summary statistics, and distribution patterns to create a baseline understanding of your data.

Time Series Plot

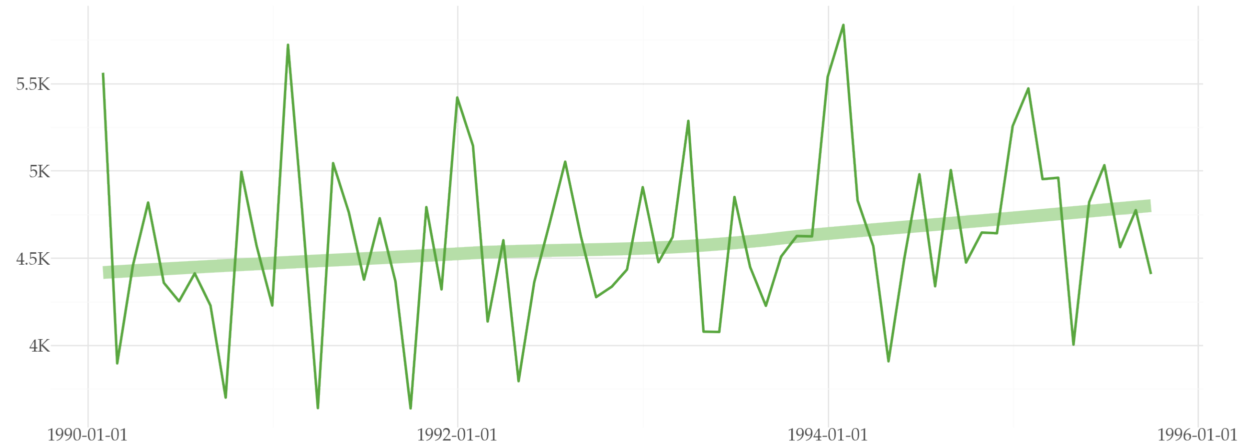


Figure 1: Time series line plot.

- A total of 69 month observations which span from January 1990 to September 1995.
- The mean value of the series is 4619.3 (median equal to 4602), with a standard deviation of 479.5. The data ranges from a minimum of 3638 to a maximum of 5836.

Data Distribution

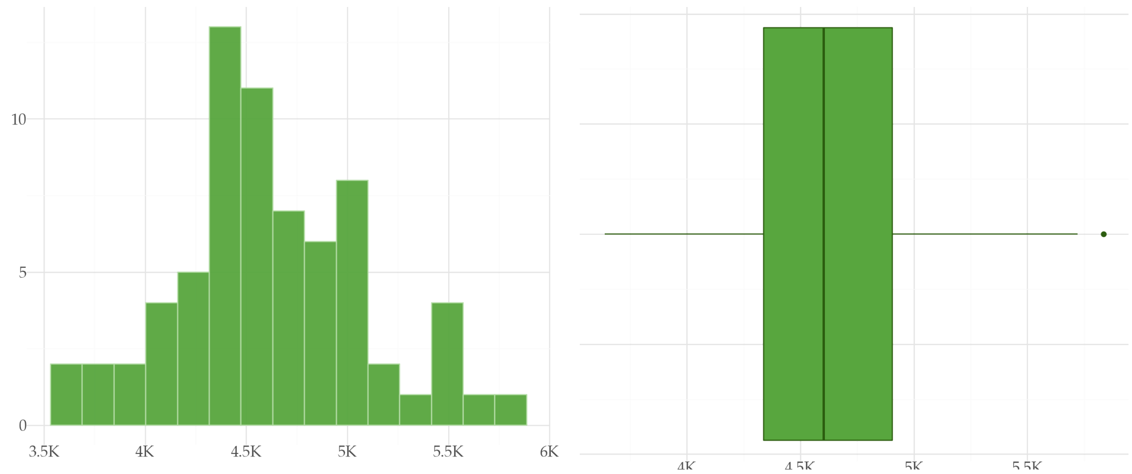


Figure 2: Distribution of the time series using an histogram (left) and a boxplot (right).

- The Kolmogorov-Smirnov test rejects the hypothesis that the series is distributed according to the following distributions: Power-law, Exponential, Pareto, and Chi-squared
- The distribution with largest p-value is Logistic (p-value equal to 0.97). But, we cannot reject the hypothesis that the data follows the following distributions

(ordered decreasingly by p-value): Exponentially Modified Gaussian distribution, Log-Normal, Gamma, Gaussian, and Cauchy.

- There are 1 outliers in the data, all of which are upper outliers. The outliers represent 1.45% of the complete data set.
- The skewness is equal to 0.28, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.
- The excess kurtosis is equal to 0.05. This value is similar to that found from data following a Gaussian distribution

Trend and Seasonality

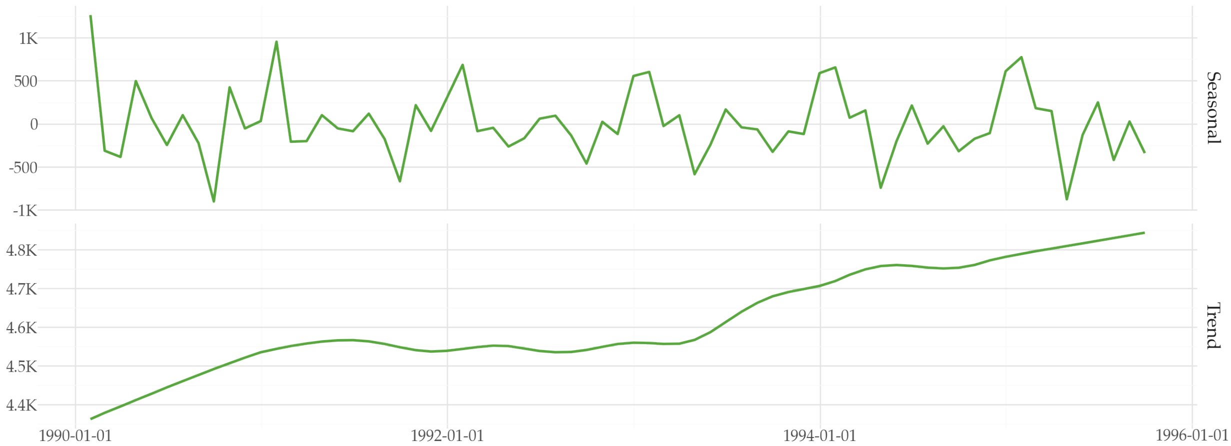


Figure 3: Seasonal and Trend components after decomposition using the STL (Season-Trend decomposition using LOESS) method.

- All hypothesis tests carried out (KPSS, Augmented Dickey-Fuller, and Philips-Perron) indicate that the time series is stationary in trend/level.
- All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is not stationary in seasonality for the specified period.

Auto-Correlation

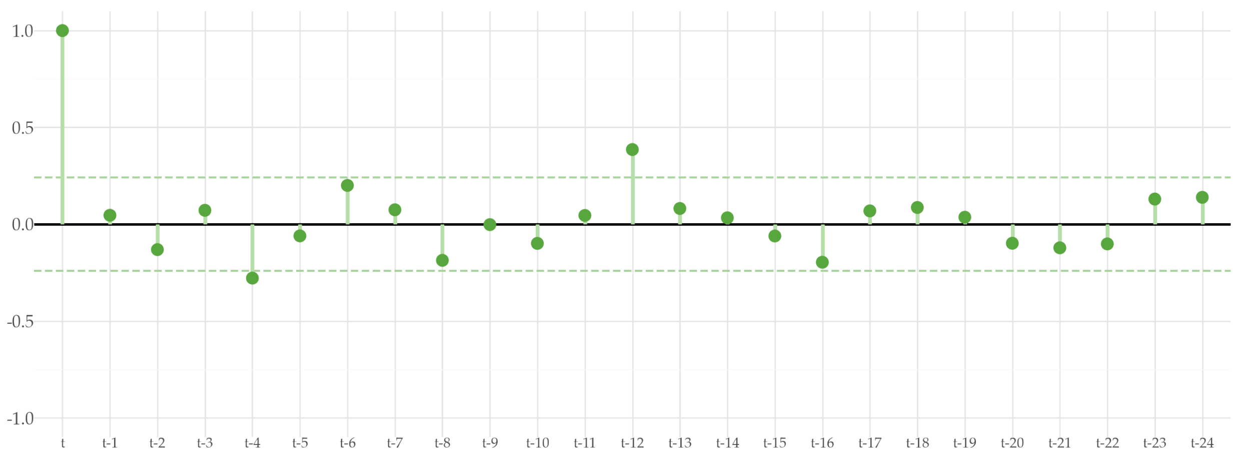


Figure 4: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-4 and t-12. Some lags show a significant negative autocorrelation (t-4), while others have a positive autocorrelation (t-12)
- Only the following lags relative to the seasonal period show a significant autocorrelation: t-12

Partial Auto-Correlation

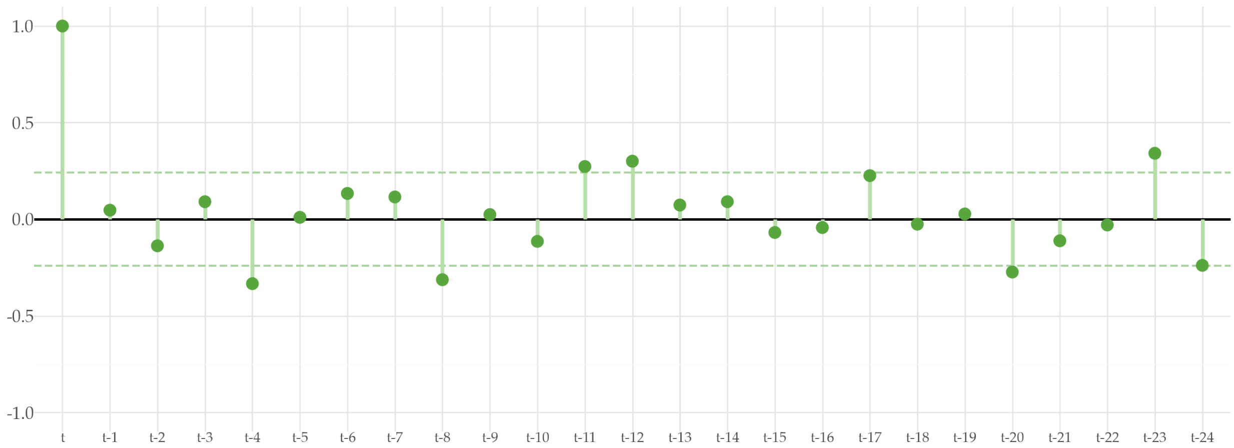


Figure 5: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-4, t-8, t-11, t-12, t-20, and t-23.
- Only the following lags relative to the seasonal period show a significant partial autocorrelation: t-12

Trend

Trend refers to the long-term change in the mean level of a time series. It reflects systematic and gradual changes in the data over time. Understanding the trend is important for identifying long-term growth or decline, structural changes, and making informed modeling decisions. This section examines the characteristics of the trend of the time series.

Trend Line Plot

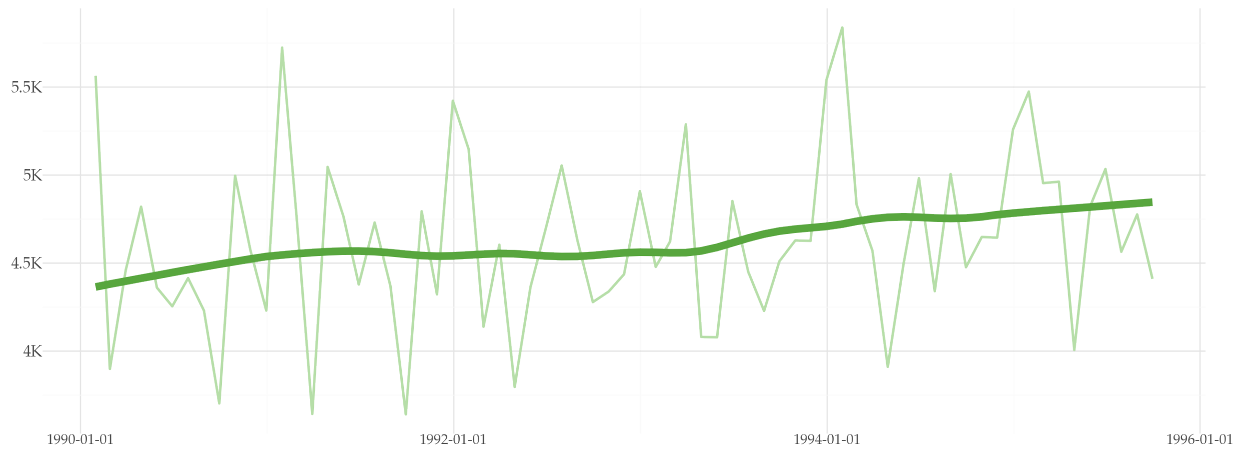
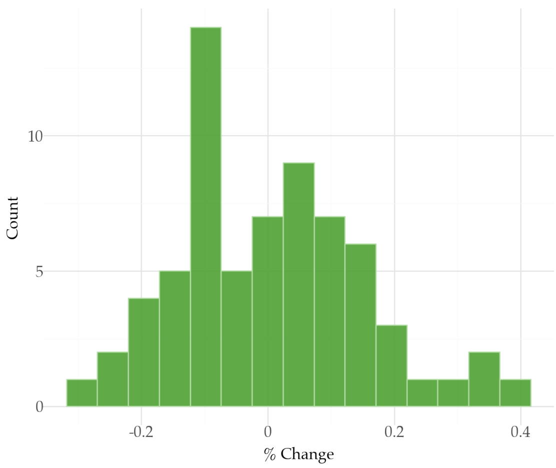


Figure 6: Time series trend plot.

- The time series has non-stationary trend according to the statistical test(s): . There is a slight upward trend.On the other hand, the method s KPSS, Augmented Dickey-Fuller, and Philips-Perron did not find evidence for the presence of trend.
- The same tests were applied to analyse whether the time series is stationary around a constant level. The method(s) Augmented Dickey-Fuller reject this hypothesis. But, the test(s) KPSS and Philips-Perron fail to reject.
- Including a trend explanatory variable which denotes the position of each observation does not improve forecasting performance.

Distribution of Differences



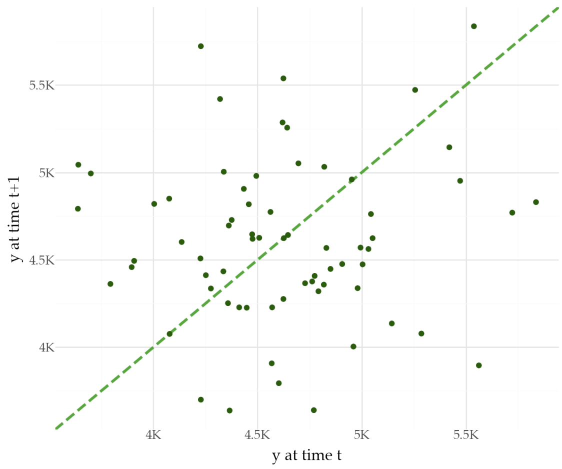


Figure 7: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The lag-plot depicts the randomness in the data. The time series shows greater randomness as the points deviate from the dotted line.

- The Kolmogorov-Smirnov test rejects the hypothesis that the differenced series is distributed according to the following distributions: Power-law, Exponential, Pareto, Log-Normal, and Chi-squared.

- The distribution with largest p-value is Gamma (p-value equal to 0.89). But, we cannot reject the hypothesis that the differenced series follows the following distributions (ordered decreasingly by p-value): Exponentially Modified Gaussian distribution, Gaussian, Logistic, and Cauchy.
- The skewness of the differenced series is equal to 0.1, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.
- The excess kurtosis of the differenced series is equal to -0.19. This value is similar to that found from data following a Gaussian distribution
- **Forecasting experiments**: Taking first differences does not improve forecasting performance.

Seasonality

Seasonality represents recurring patterns or cycles that appear at regular intervals in time series data. These are predictable fluctuations that reflect periodic influences such as monthly, quarterly, or yearly cycles. Understanding seasonal patterns is crucial for forecasting, trend analysis, and identifying anomalies. This section examines the presence, strength, and characteristics of seasonal components in the input time series.

Seasonal Line Plot (Monthly)

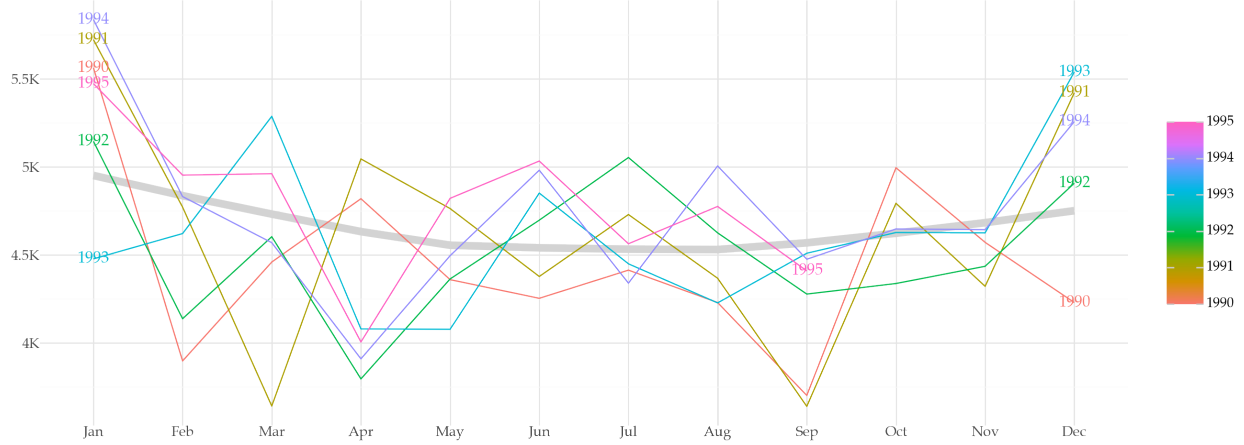


Figure 8: Seasonal plot of monthly values grouped by year.

- All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is not stationary in seasonality for a yearly period.
- **Forecasting experiments**: Including monthly information in the predictive model decreases forecasting performance. This information was included as Fourier terms and repeating basis function terms in the explanatory variables.

Seasonal Sub-series Plot (Monthly)

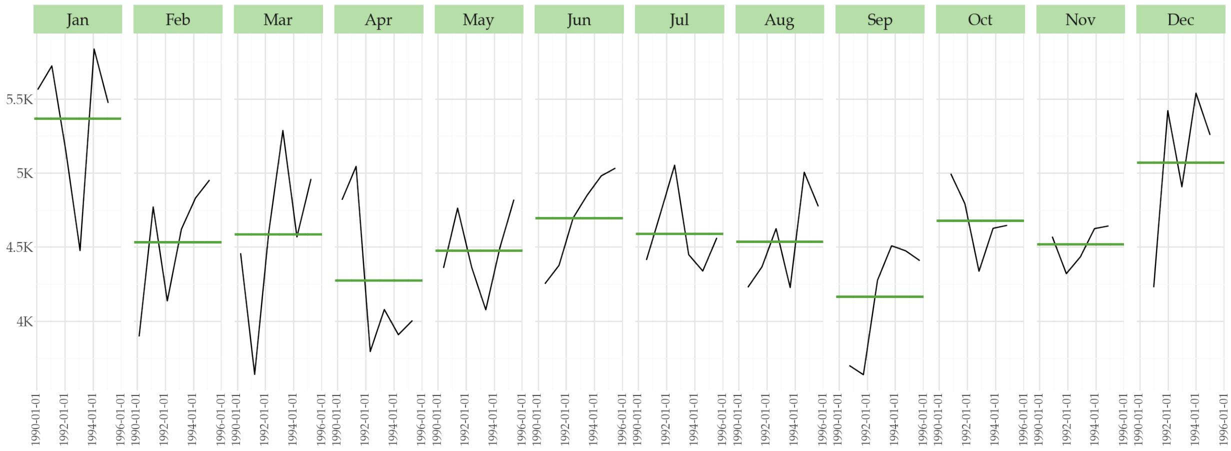


Figure 9: Monthly seasonal sub-series. This plot helps to understand how the data varies within and across monthly groups.

- Overall, there is a slight evidence that the time series is not stationary around a constant level. But, the data is constant around a level in each Month.
- **Forecasting experiments**: There is evidence for a yearly seasonal pattern from statistical tests. Yet, including information about this period in the forecasting model decreased its performance.

Mean and Standard Deviation Analysis (Monthly)

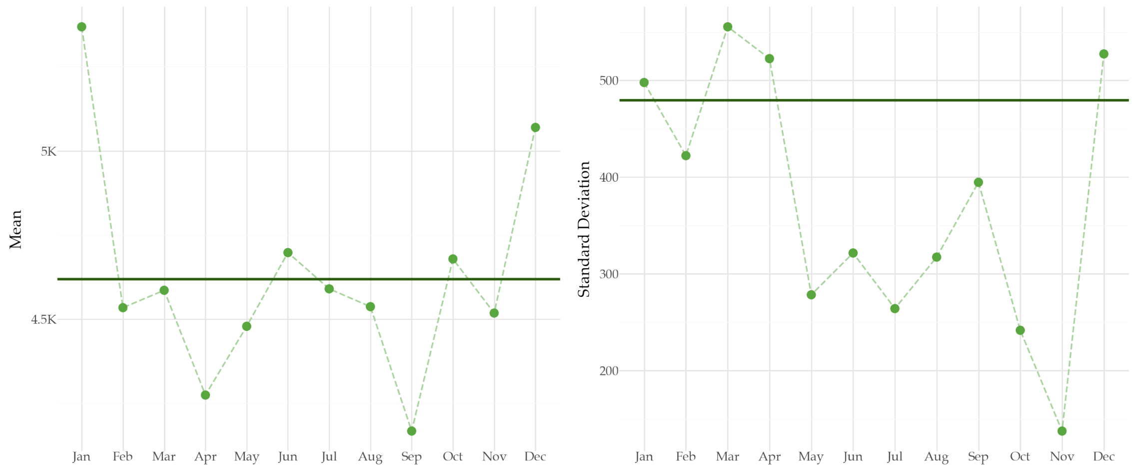


Figure 10: Mean plot (left) and standard deviation plot (right) of monthly values.

- The data shows significantly different mean when grouped by month.

Seasonal Sub-series Plot (Quarterly)

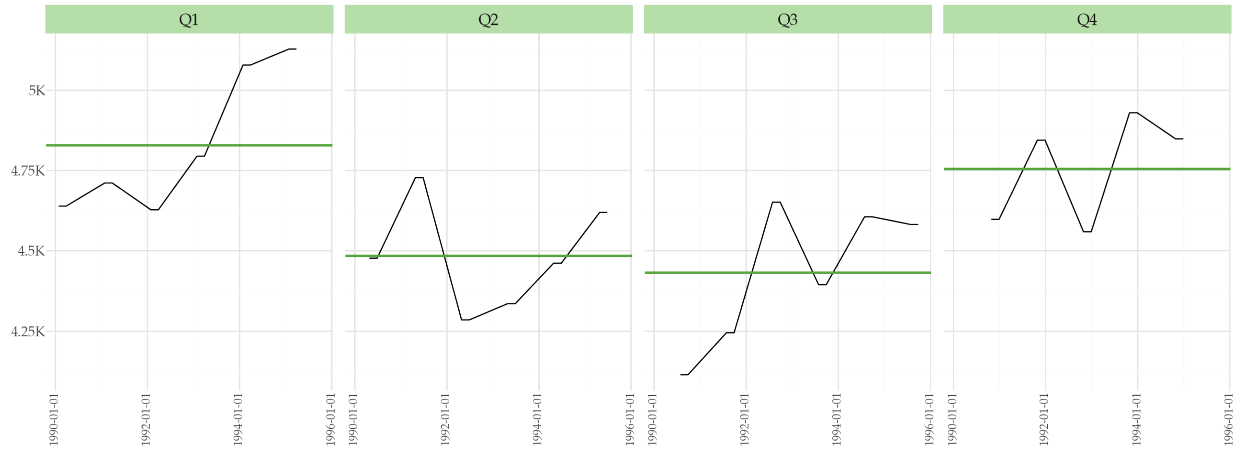


Figure 11: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- The following tests indicate that the time series is non-stationary in seasonality for a Quarterly period: OCSB. On the other hand, other tests (Wang-Smith-Hyndman) fail to reject the stationary null hypothesis.

- Overall, there is a slight evidence that the time series is not stationary around a constant level. But, the data is constant around a level in each Quarter.

- **Forecasting experiments**: There is evidence for a quarterly seasonal pattern from statistical tests. Yet, including information about this period in the forecasting model decreased its performance.

Mean and Standard Deviation Analysis (Quarterly)

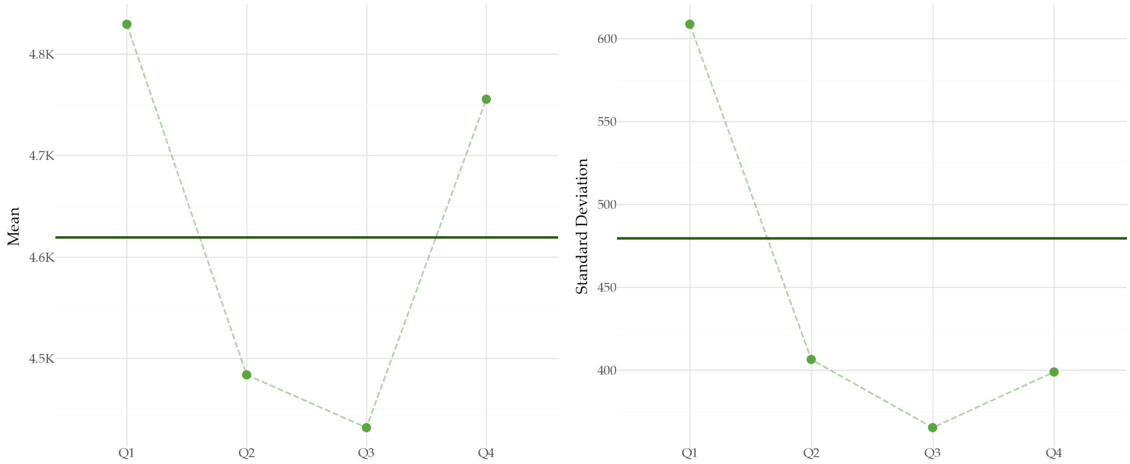


Figure 12: Mean plot (left) and standard deviation plot (right) of quarterly values.

- The data shows significantly different mean when grouped by quarter.