

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: 2024-11-03 22:48

Series Name: M1

Table of Contents

1	Data Overview	Time series fundamental characteristics and statistical properties
2	Trend	Long-term time series growth and dynamics. Analysis of level stabilization methods.
3	Seasonality	Analysing recurring patterns in the time series. Assessing the impact of different seasonality modeling strategies
4	Variance	Exploring the variability of values over time. Assessing the impact of variance stabilization methods

Other aspects were explored but omitted from the final report:

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

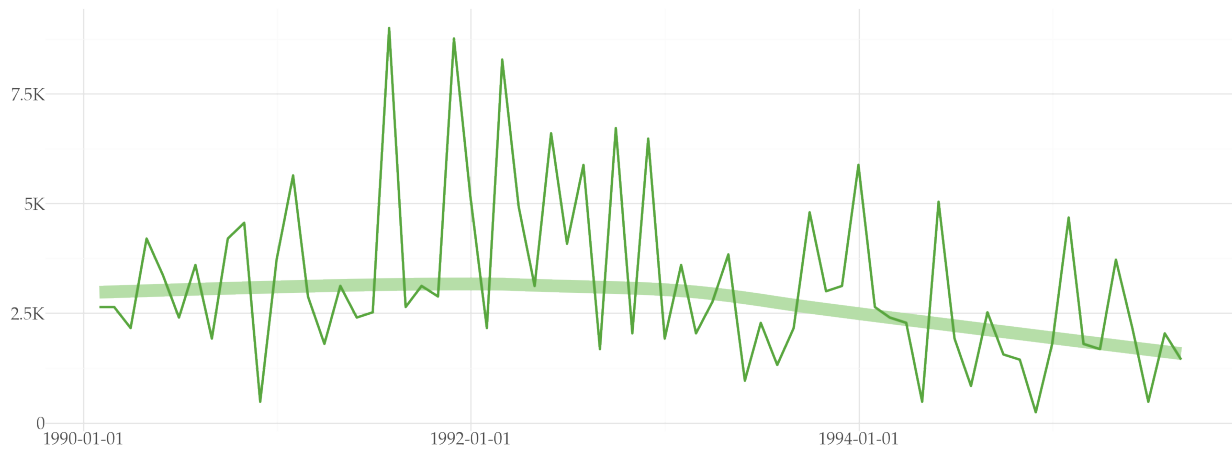


Figure 1: Time series line plot.

- A total of 68 observations which span from January 1990 to August 1995. These are collected with a monthly sampling frequency.
- The data ranges from a minimum of 240 to a maximum of 9000, starting in 2640 and ending in 1440 during the observed period. The average growth percentage per observation is 57.75% (median growth equal to -7.69%), with an average value of 3185.29. There are no missing values in the time series.

Trend, Seasonality, and Residuals

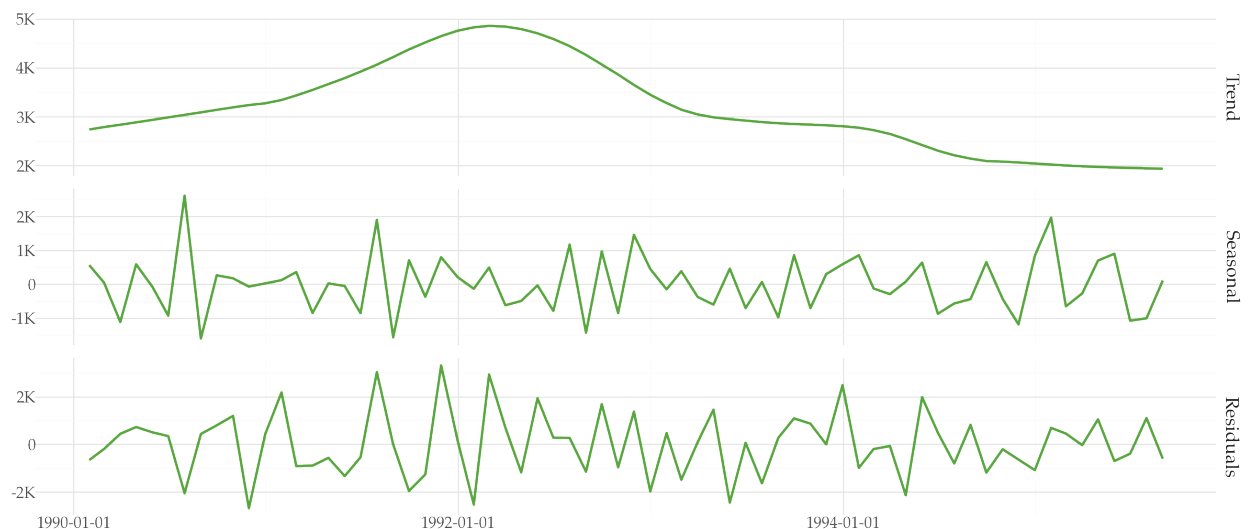


Figure 2: Seasonal, Trend, and Residuals components after decomposition on a monthly frequency using the STL (Season-Trend decomposition using LOESS) method.

- The trend strength (ranges from 0 to 1) is 0.35. The following tests indicate that the time series is non-stationary in trend or level: KPSS. On the other hand, other tests

(Augmented Dickey-Fuller and Philips-Perron) fail to reject the hypothesis that the data is stationary.

- The seasonal strength (also ranges from 0 to 1) is 0.34. The following tests indicate that the time series is non-stationary in seasonality for the specified period: OCSB. On the other hand, other tests (Wang-Smith-Hyndman) fail to reject stationarity.
- The STL decomposition residuals show balanced behavior: 50.0% of residuals are positive and 50.0% negative. The average magnitude of positive residuals is 1060.918 compared to -1046.387 for negative residuals. In terms of auto-correlation structure, the residuals show significant temporal dependency in some of the first 12 lags according to the Ljung-Box test. This suggests that the decomposition method is missing some systematic patterns.

Auto-Correlation

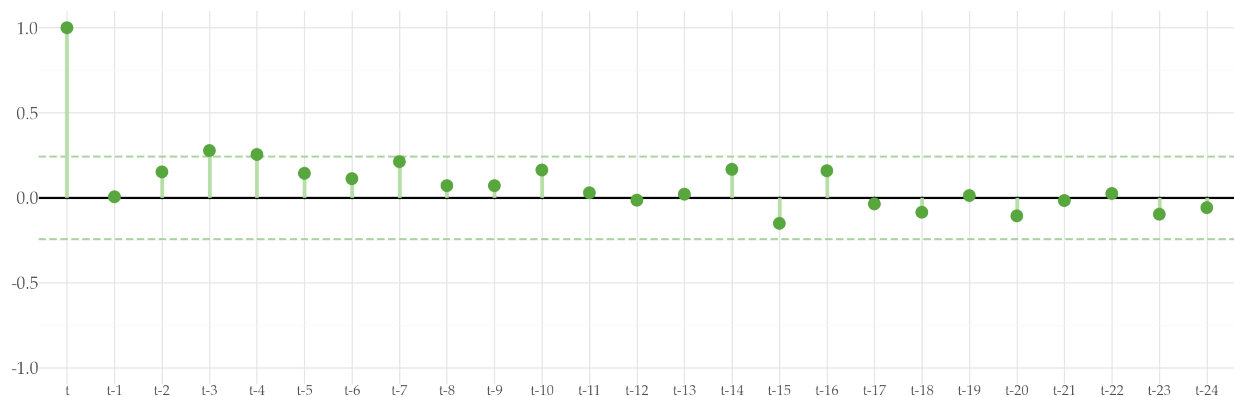


Figure 3: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-3 and t-4. The autocorrelation is positive for all lags with a significant value.
- None of the lags relative to the seasonal period (t-12 and t-24) show any significant autocorrelation.

Partial Auto-Correlation

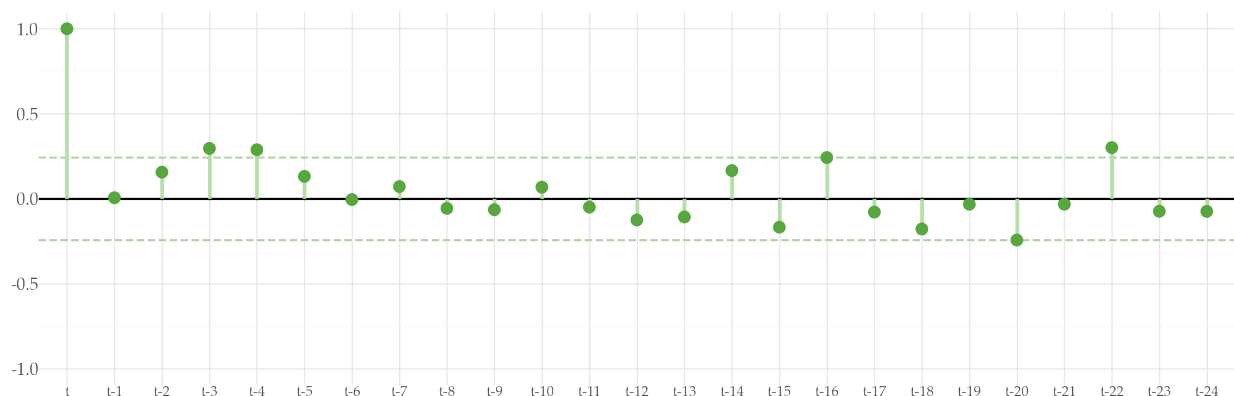


Figure 4: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-3, t-4, t-20, and t-22.
- None of the lags relative to the seasonal period (t-12 and t-24) show any significant partial autocorrelation.

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

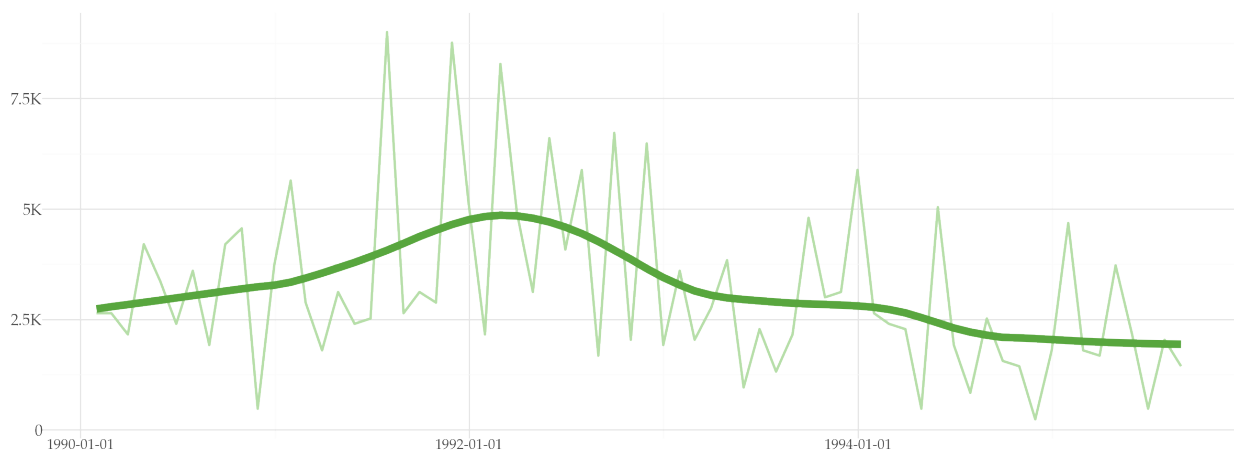


Figure 5: Time series trend plot.

- There is a slight downward trend. This trend is non-stationary (i.e. not deterministic) according to the statistical test(s): KPSS. The test(s) Augmented Dickey-Fuller and Philips-Perron did not find evidence for non-stationarity around a deterministic trend.
- The same tests were applied to analyse whether the time series is stationary around a constant level. The method(s) KPSS and Augmented Dickey-Fuller reject this hypothesis. But, the test(s) Philips-Perron fail to reject stationarity.
- **Preliminary experiments:** Including a trend explanatory variable which denotes the position (row id) of each observation improves forecasting accuracy. These experiments were conducted using a LightGBM algorithm and evaluated using SMAPE loss function. Using only lag-based features the model achieved a SMAPE of 79.33% on the test set. Including the trend variable improved the SMAPE to 64.67%.

Distribution of Differences

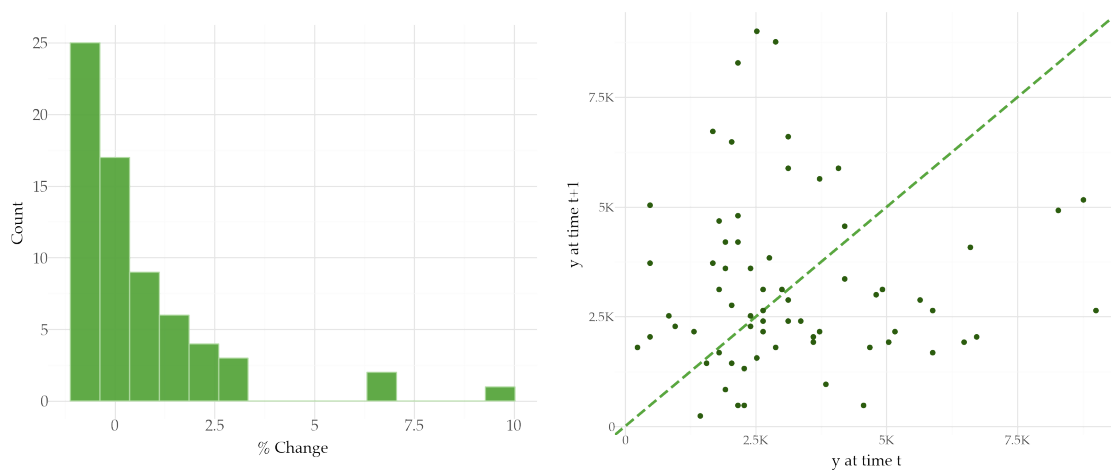


Figure 6: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The lag-plot depicts the randomness in the data. The time series shows greater randomness 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 Logistic (p-value equal to 1.0). But, we cannot reject the hypothesis that the differenced series follows the following distributions (ordered decreasingly by p-value): Exponentially Modified Gaussian distribution, Gamma, Gaussian, and Cauchy.
- The skewness of the differenced series is equal to 0.26, 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 -0.14. 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)

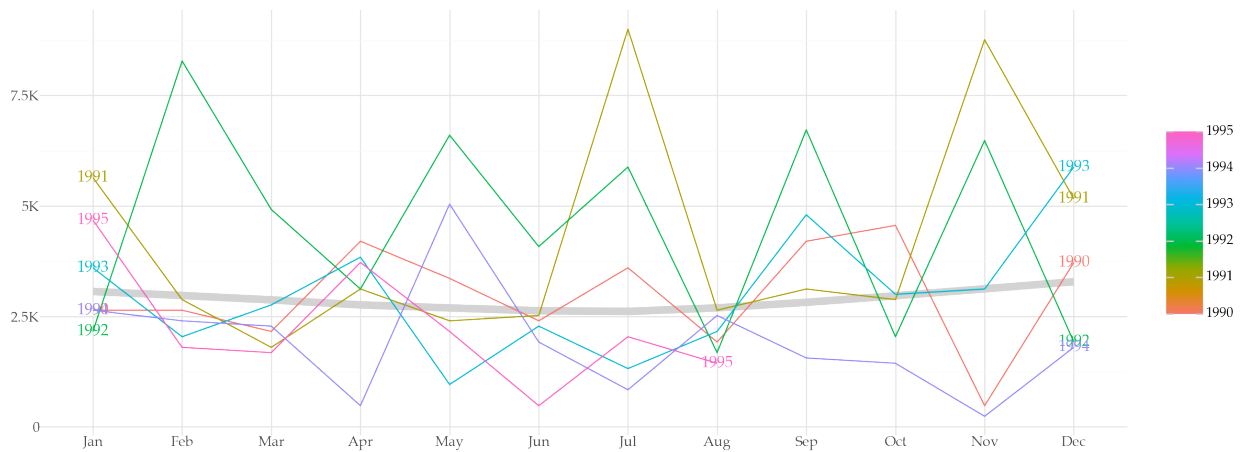


Figure 7: Seasonal plot of monthly values grouped by year.

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

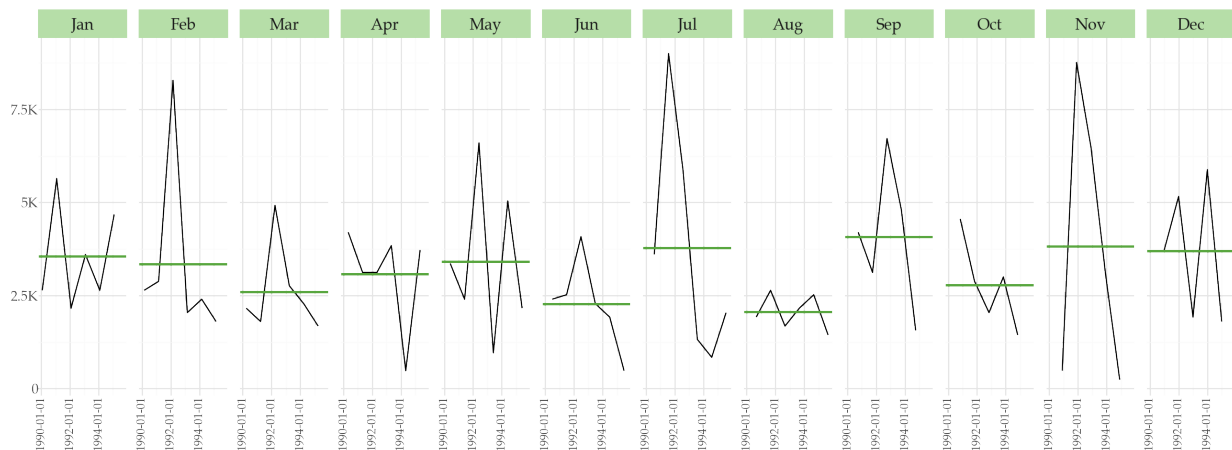


Figure 8: Monthly seasonal sub-series. This plot helps to understand how the data varies within and across monthly groups.

- Statistical tests were carried out to check for differences among means and variances across months. No significant differences were found.
- Overall, there is a reasonable 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.

Seasonal Sub-series Plot (Quarterly)

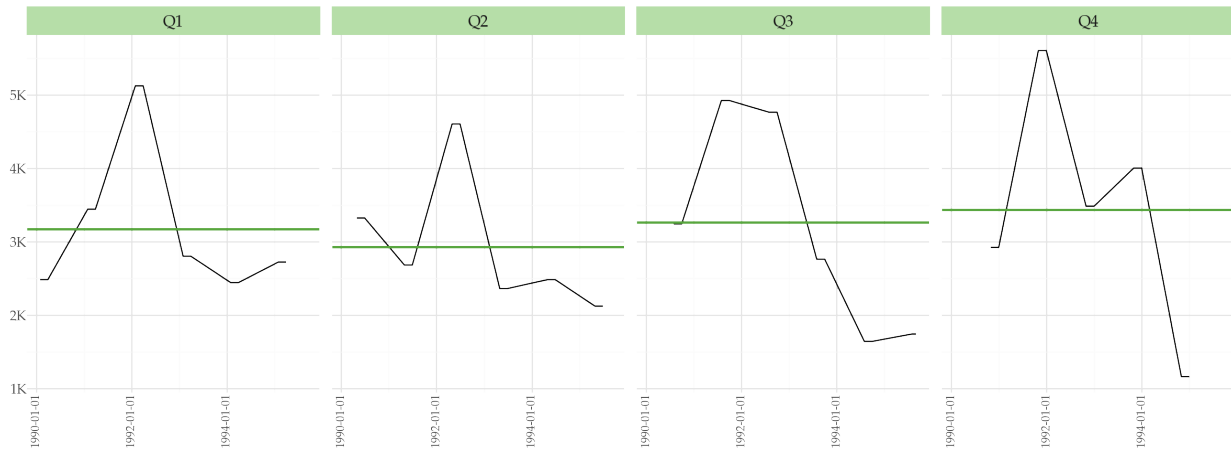


Figure 9: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- Statistical tests were carried out to check for differences among means and variances across quarters. No significant differences were found.
- 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 reasonable 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. Besides, including information about this period in the forecasting model improved its performance.

Variance

Variance measures how data points spread around the average value in your time series. This section examines whether the variability remains stable (homoskedastic) or changes (heteroskedastic) over time. Understanding variance patterns is crucial for selecting appropriate modeling techniques, which can have a significant impact on forecasting accuracy.

Heteroskedasticity Testing

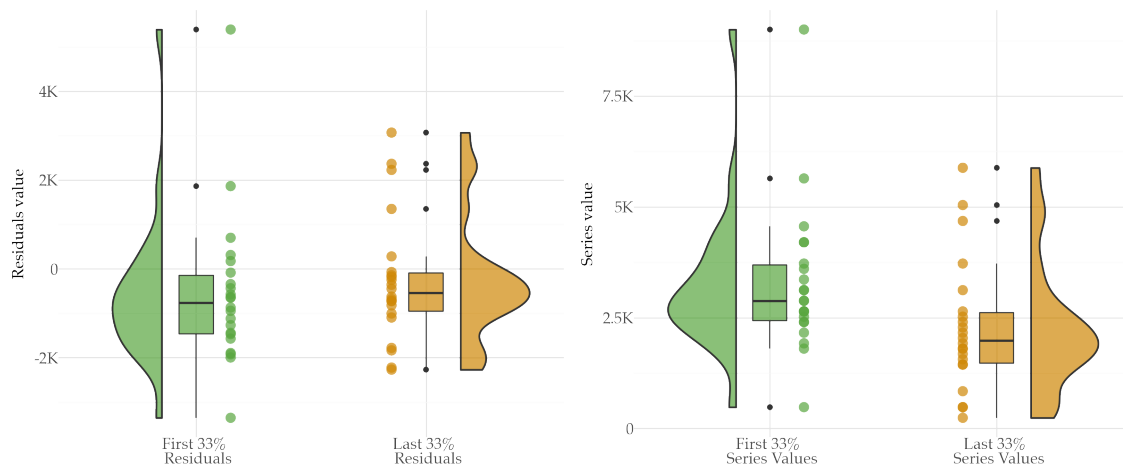


Figure 10: Time series residuals analysis. Difference in the distribution of the residuals (left) and the series (right) in the first and last thirds of the series, following a Goldfeld-Quandt partition.

- In the analysis of seasonality we did not find significant differences in the dispersion among periodic groups of observations.
- No statistical evidence was found for the hypothesis that the time series is heteroskedastic, according to the White, Breusch-Pagan, and Goldfeld-Quandt tests.
- ****Forecasting experiments****: Transforming the series with either the logarithm or the Box-Cox method improved forecasting performance.
- In the original scale, the Logarithm distribution is a reasonable fit to the data. But, after taking the Exponentially Modified Gaussian distribution transformation, a Logistic distribution was found to be a better fit.