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: 2025-02-04 16:02

Series Name: M1080

Table of Contents

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

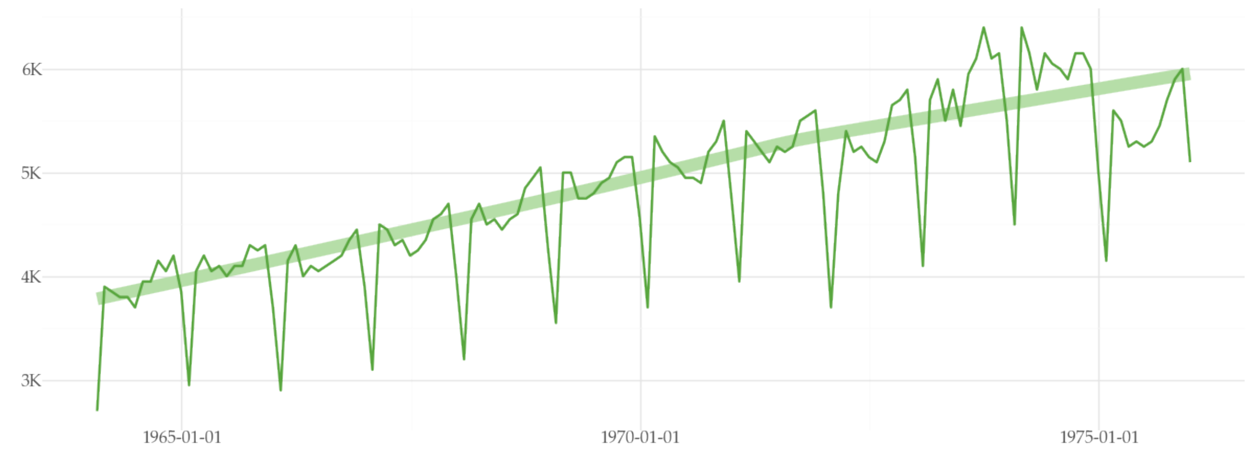

Figure 1: Time series line plot.

- A total of 144 observations spanning from January 1964 to December 1975. These are collected with a monthly sampling frequency.
- The data ranges from a minimum of 2700 to a maximum of 6400, starting in 2700 and ending in 5100 during the observed period. The average growth percentage per observation is 1.28% (median equal to 0.88%), with an average value of 4815.97. 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 is 0.98 (ranges from 0 to 1). All hypothesis tests carried out (KPSS and Philips-Perron) indicate that the time series is stationary in trend or level.

- The seasonal strength is 0.95 (ranges from 0 to 1). All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is not stationary in seasonality for the specified period.
- The STL decomposition residuals show unbalanced behavior: 55.56% of residuals are positive and 44.44% negative. The average magnitude of positive residuals is 58.616 compared to -68.325 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

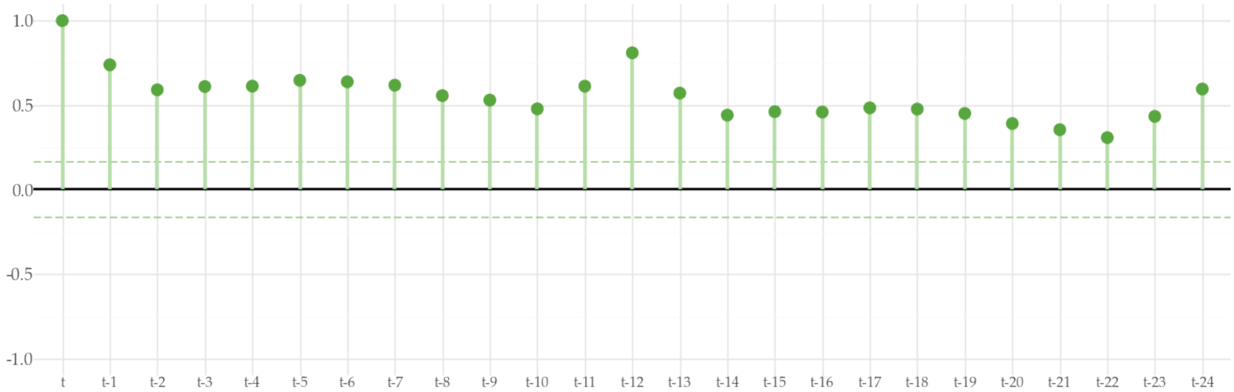

Figure 3: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-1, t-2, t-3, t-4, t-5, t-6, t-7, t-8, t-9, t-10, t-11, t-12, t-13, t-14, t-15, t-16, t-17, t-18, t-19, t-20, t-21, t-22, t-23, and t-24. The autocorrelation is positive for all lags with a significant value.
- All lags relative to the seasonal period (t-12 and t-24) show a significant autocorrelation.

Partial Auto-Correlation

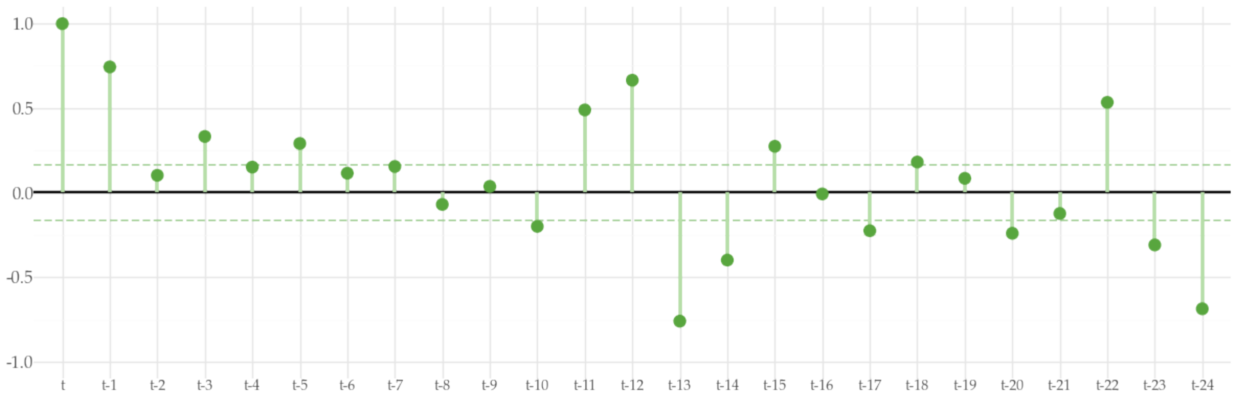

Figure 4: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-3, t-5, t-10, t-11, t-12, t-13, t-14, t-15, t-17, t-18, t-20, t-22, t-23, and t-24.
- All lags relative to the seasonal period (t-12 and t-24) show a 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

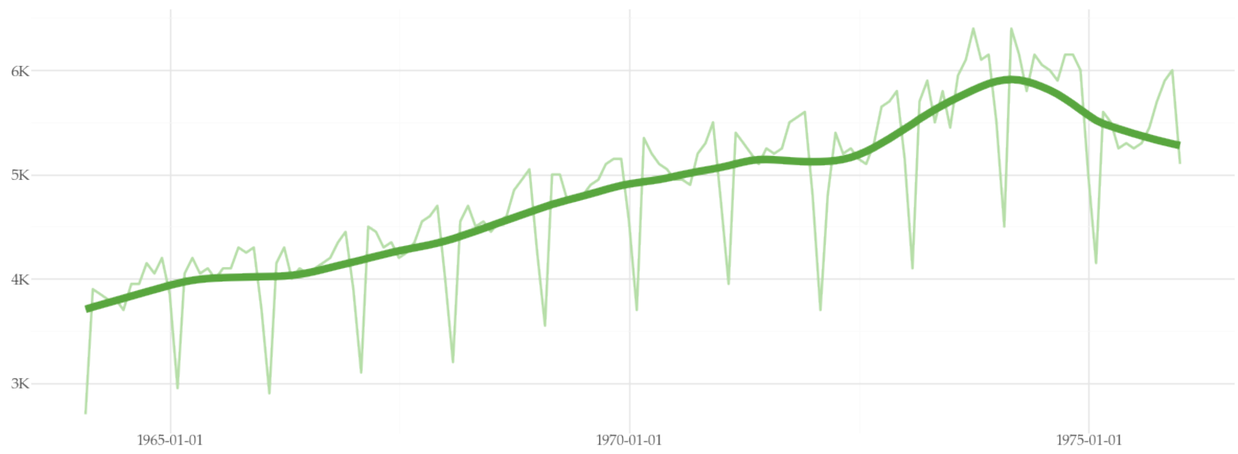

Figure 5: Time series trend plot.

- There is a strong upward trend. The tests KPSS and Philips-Perron did not find evidence for non-stationarity around a deterministic trend.
- The same tests were applied to analyse stationarity around a constant level. All tests failed to reject this hypothesis.
- **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 9.1% on the test set. Including the trend variable improved the SMAPE to 7.77%.

Long-term Growth

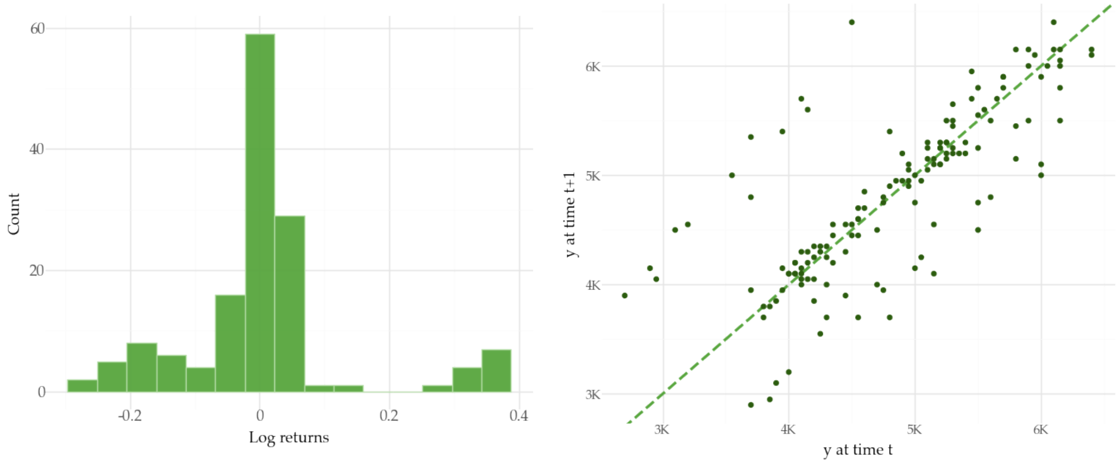

Figure 6: Distribution of log differences (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram show the distribution of these changes using log returns. The lag-plot depicts the randomness in the data. The time series shows greater randomness as the points deviate from the dotted line.

- The time series has an average growth (log returns) of 0.004 (median equal to 0.01). The volatility of the returns in terms of standard deviation is 0.13. The skewness of the log differenced series is equal to 0.98, indicates that the right tail is long relative to the left tail. The excess kurtosis of the log differenced series is equal to 2.26. This indicates a heavy tailed distribution.

- Concerning the symmetry of returns, 51.05% of the log differences are positive. The average of positive returns is 0.08, while the average of negative returns (43.36% of all returns) is -0.08. Overall, there are 75 return direction changes (52.82% of the data points)

- In the tails, 9.79% of returns fall beyond 2 standard deviations from the mean. The largest positive return is 0.37 on January 1967. Conversely, the largest decline is -0.27 (on December 1964).

- **Preliminary experiments:** Modeling the time series of first differences may improve forecasting accuracy. Experiments were conducted using a LightGBM algorithm and evaluated using SMAPE loss function. Using the original time series led to a 9.1% SMAPE. The scores using the differenced and log differenced time series are 4.24% and 3.82%, respectively.

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)

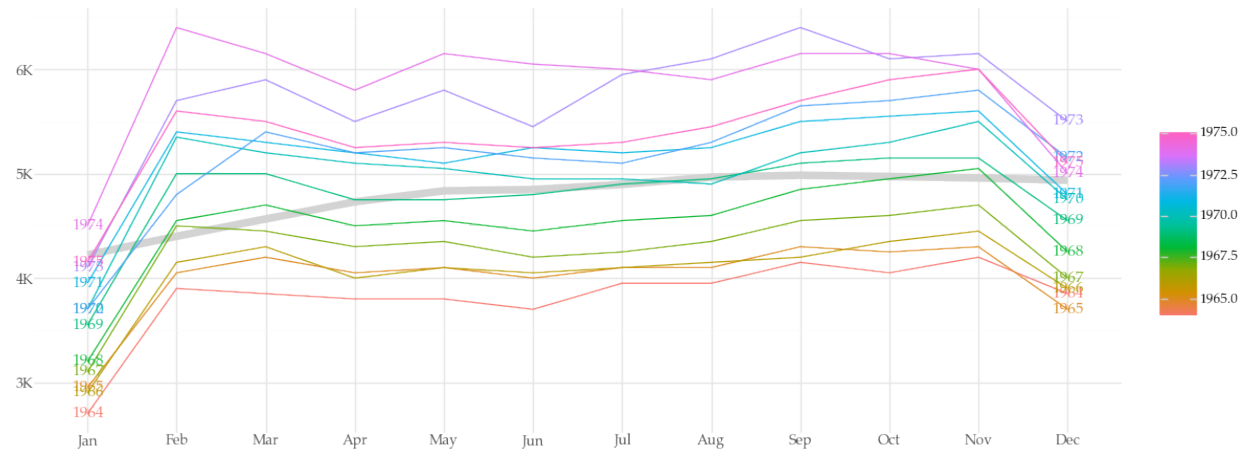

Figure 7: Seasonal plot of monthly values grouped by year.

- The seasonal strength is 0.95. This score ranges from 0 to 1 and values above 0.64 are considered significant. All hypothesis tests carried out (Wang-Smith-Hyndman and OCSB) indicate that the time series is not stationary in seasonality for a yearly period.

- **Preliminary experiments:** Modeling yearly patterns does not improve forecast accuracy. Different approaches were tested relative to a base model using only lag-based features (9.1% SMAPE):
  - Fourier terms: 9.45% SMAPE
  - Seasonal differencing: 11.34% SMAPE
  - Monthly time features: 9.1% SMAPE

Mean and Standard Deviation Analysis (Monthly)

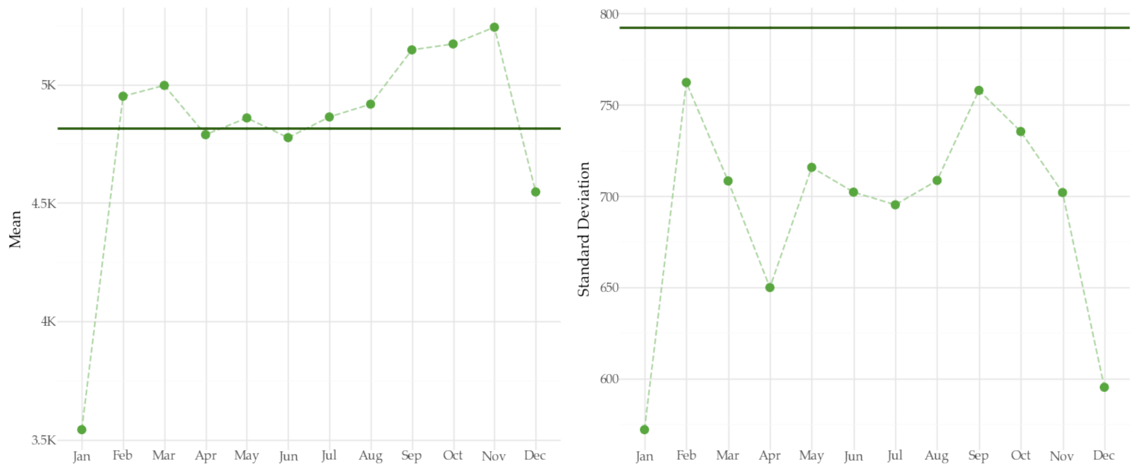

Figure 8: Mean plot (left) and standard deviation plot (right) of monthly values. The horizontal line shows the overall value across groups, while the dots show the value in the corresponding group.

- Analysis of monthly patterns shows significant differences in central tendency (Kruskal-Wallis test) but similar dispersion (Levene's test).

Seasonal Sub-series Plot (Quarterly)

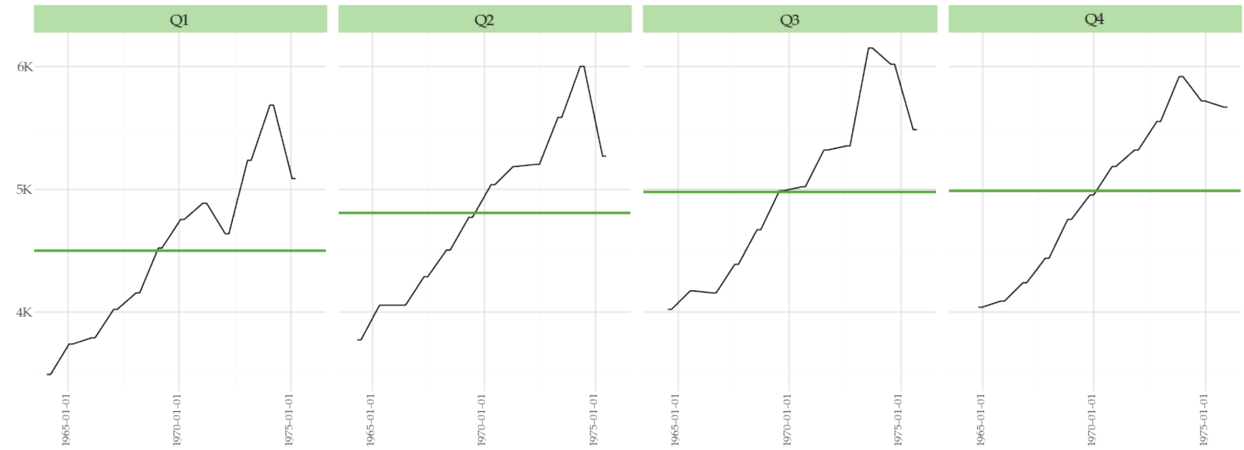

Figure 9: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- Tests for quarterly seasonal stationarity show mixed results: the OCSB test indicates presence of a seasonal unit root, while the Wang-Smith-Hyndman test suggests stationarity.
- **Preliminary experiments:** There is evidence for a quarterly seasonal pattern based on statistical tests. Besides, modeling quarterly patterns can improve forecast accuracy. Different approaches were tested relative to a base model using only lag-based features (9.1% SMAPE):
    - Fourier terms: 9.78% SMAPE
    - Quarterly seasonal differencing: 8.08% SMAPE
    - Quarterly time features: 9.1% SMAPE

Mean and Standard Deviation Analysis (Quarterly)

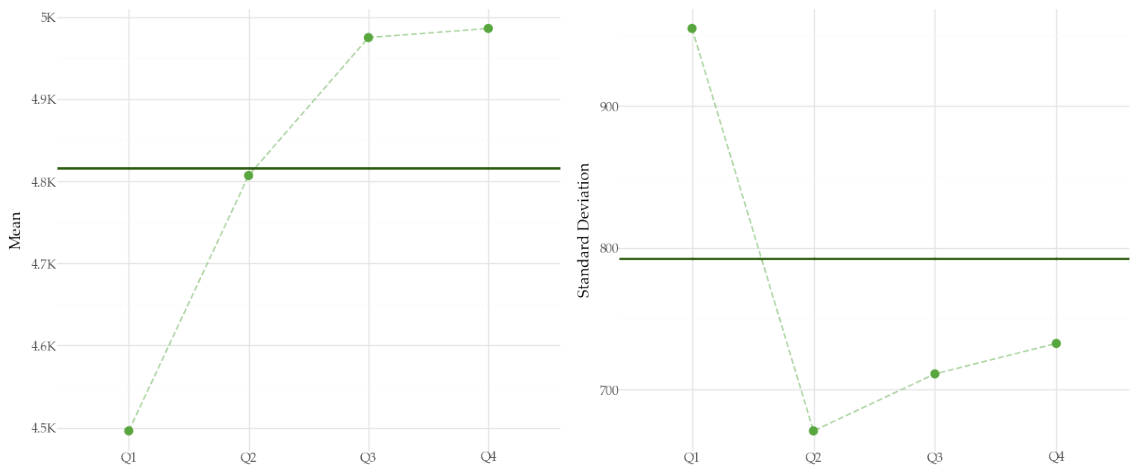

Figure 10: Mean plot (left) and standard deviation plot (right) of quarterly values. The horizontal line shows the overall value across groups, while the dots show the value in the corresponding group.

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

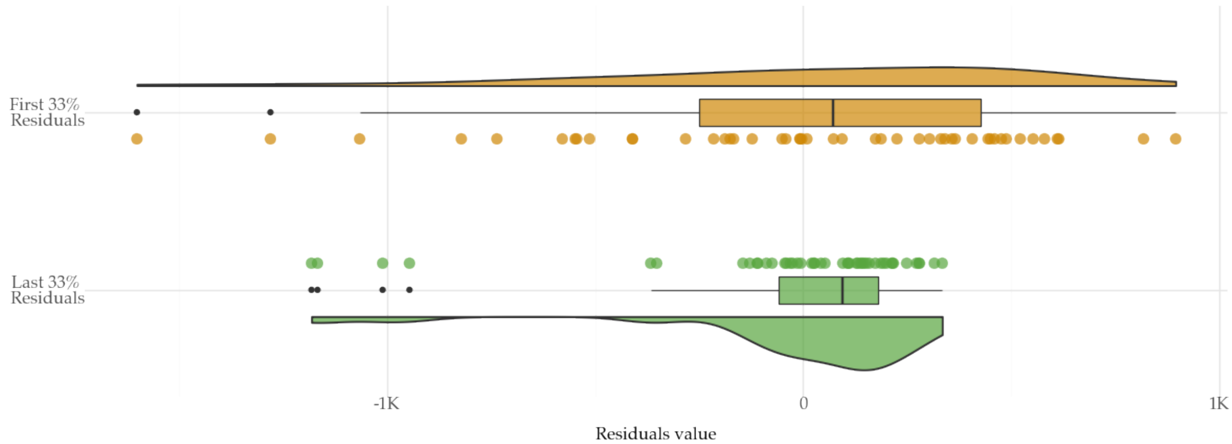

Figure 11: Time series residuals analysis based on a linear trend model. Difference in the distribution of the residuals in the first and last thirds of the series, following a Goldfeld-Quand partition.

- The following tests suggest that the time series is heteroskedastic: White. But, other tests (Breusch-Pagan and Goldfeld-Quandt) fail to reject the hypothesis that the time series has a constant variance. The residuals are based on a linear trend model.
- Variance in seasonal periods according to Levene's test
    - Quarterly groups: no differences in variance
    - Monthly groups: no differences in variance
- **Preliminary experiments:** Three variance stabilization preprocessing techniques were tested to improve the forecast accuracy of an auto-regressive LightGBM (with 9.1% SMAPE using lag-based features):
    - Log returns: 3.82% SMAPE
    - Log transformation: 9.54% SMAPE
    - Box-Cox transformation: 9.12% SMAPE

Change Detection

Change points denote significant shifts in the underlying distribution of time series. These structural changes can manifest as sudden shifts in level, trend, variance, or seasonal patterns. Detecting and understanding these points is crucial as they often indicate important events or regime changes that affect modeling decisions. This section identifies potential change points and assesses their impact on the overall analysis strategy.

Change Points

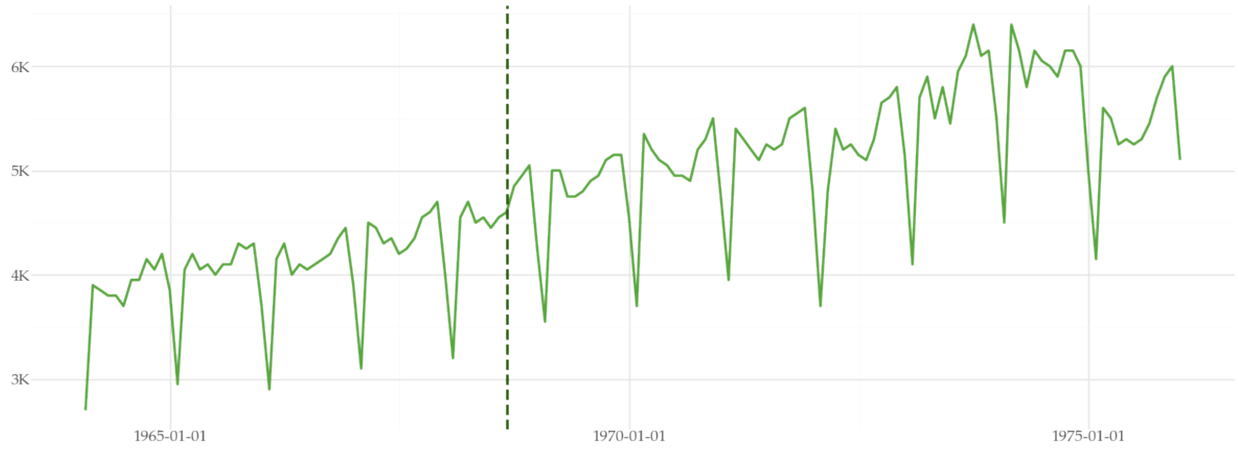

Figure 12: Time series plot with marked change points according to the PELT method.

- A single change point was found in the time series.
- The change point was found at August 1968 where the time series shows an increasing level.

Effect on Model Parameters

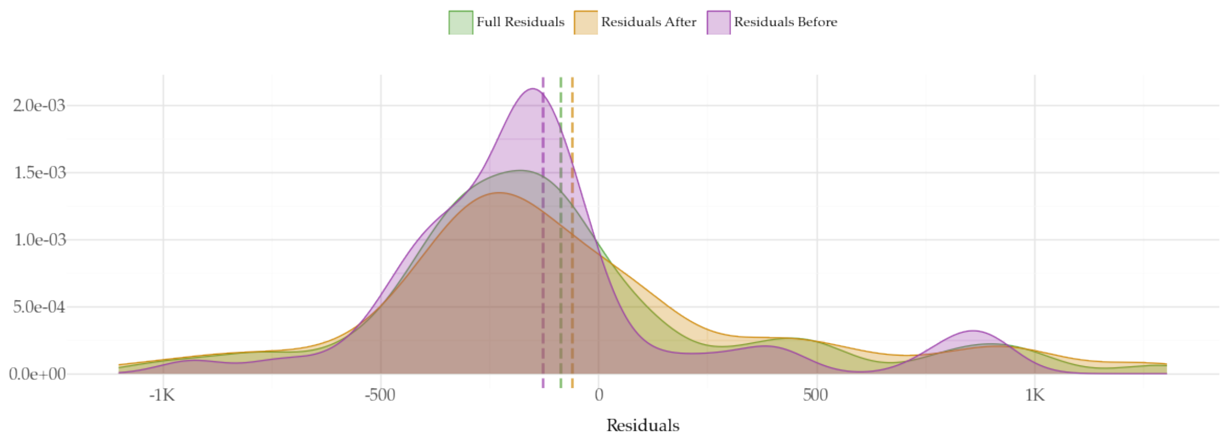

Figure 13: Distribution of the residuals of an ARIMA model before and after the first detected change point. The plot compares three kernel density estimates: residuals from the pre-change model, post-change model, and full series model. This comparison helps assess whether the structural break affects model adequacy and error distribution properties.

- A Chow test was conducted using an ARIMA(2, 1, 2) model. The test fails to reject the null hypothesis of parameter stability. This suggests that the ARIMA parameters remain stable before and after the first detected change point, suggesting the underlying process structure remained similar despite the level shift.

- **Preliminary experiments:** Adding a step intervention at the change point did not affect the model performance. The baseline SMAPE of 9.1% remained the same when including the intervention (9.1%).