

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: 2024-11-11 15:26

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
- 5 Change Detection**
Change detection in the time series distribution

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

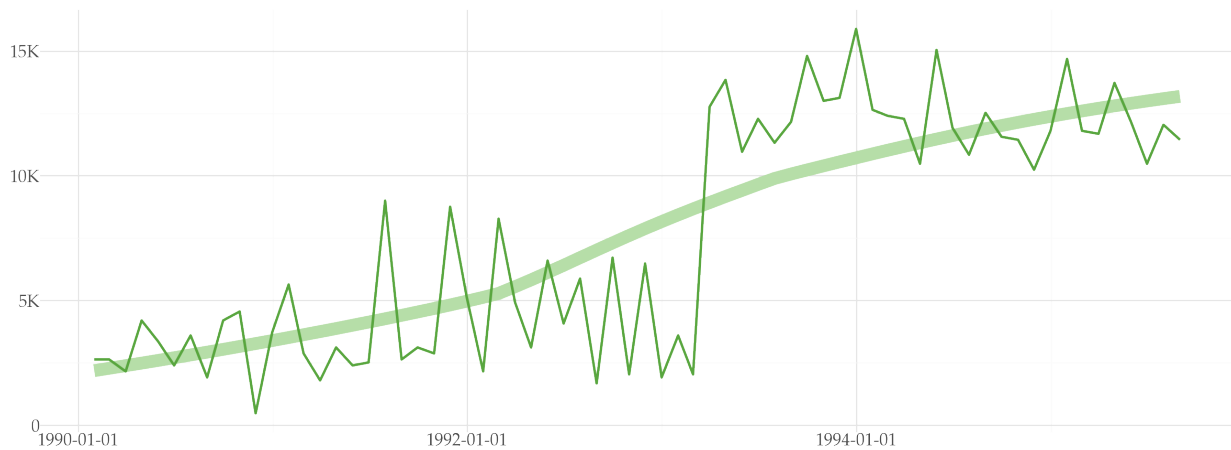


Figure 1: Time series line plot.

- 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 480 to a maximum of 15880, starting in 2640 and ending in 11440 during the observed period. The average growth percentage per observation is 34.18% (median growth equal to -1.04%), with an average value of 7597.06. There are no missing values in the time series.

Trend, Seasonality, and Residuals

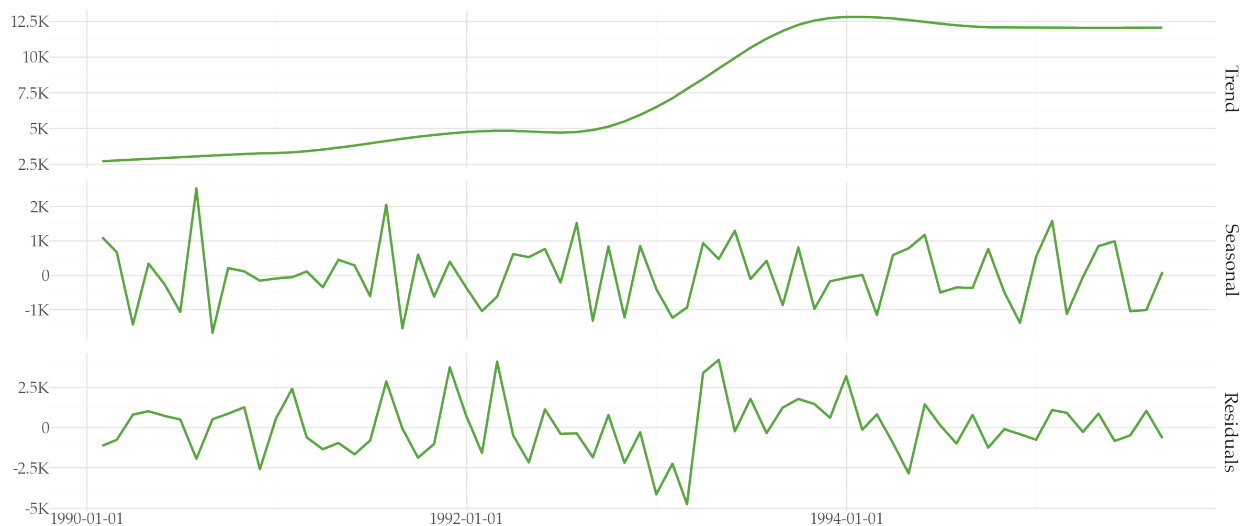


Figure 2: Seasonal, Trend, and Residuals components after decomposition on a monthly frequency using the STL (Season-Trend decomposition using LOESS) method.

- The trend strength (ranges from 0 to 1) is 0.85. The following tests indicate that the time series is non-stationary in trend or level: Augmented Dickey-Fuller. On the other hand, other tests (KPSS and Philips-Perron) fail to reject the hypothesis that the data is stationary.

- The seasonal strength (also ranges from 0 to 1) is 0.27. 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 balanced behavior: 46.0% of residuals are positive and 54.0% negative. The average magnitude of positive residuals is 1498.638 compared to -1241.273 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

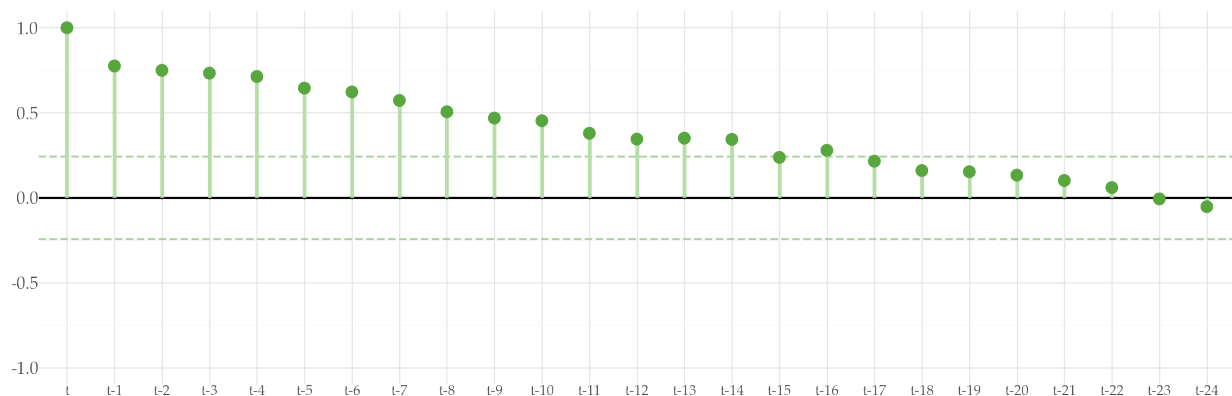


Figure 3: Auto-correlation plot up to 24 lags.

- 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, and t-16. 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

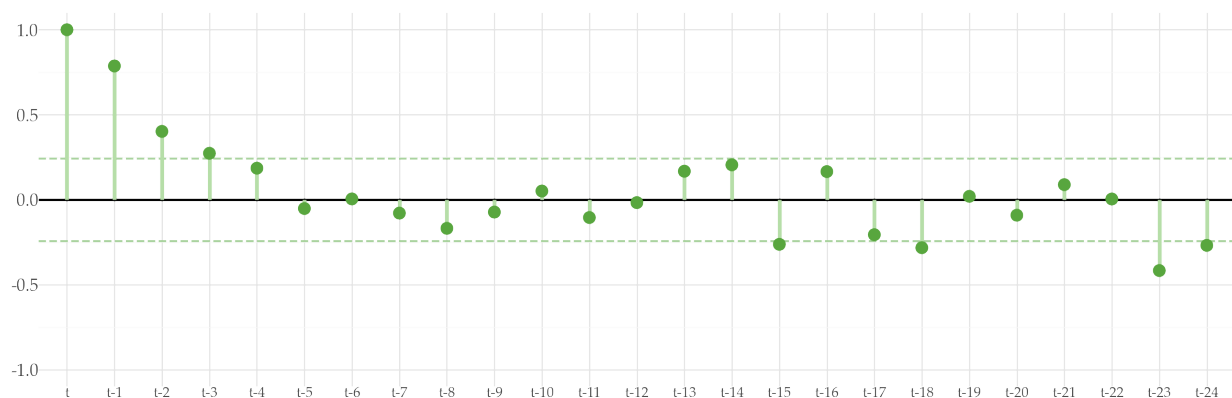


Figure 4: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-2, t-3, t-15, t-18, t-23, and t-24.
- 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

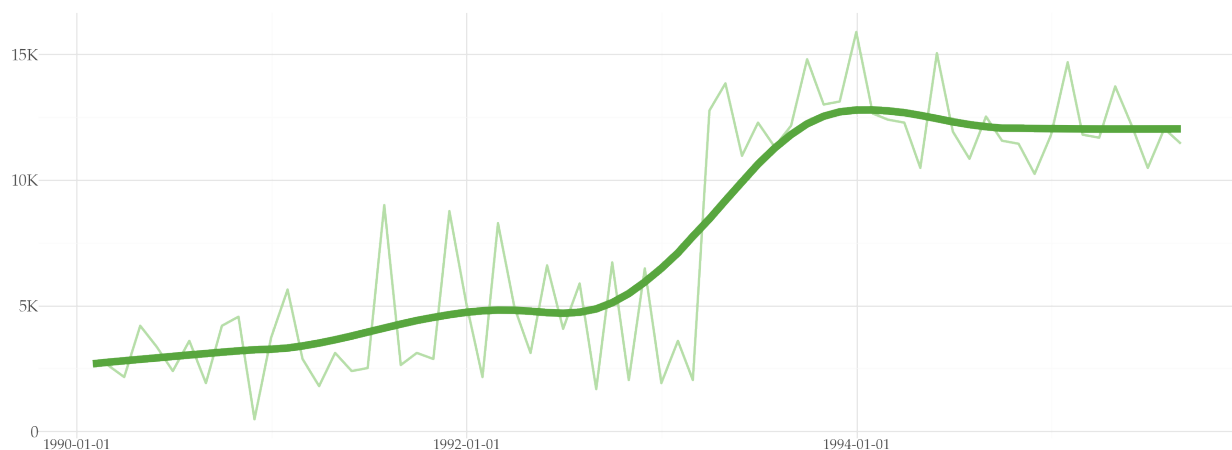


Figure 5: Time series trend plot.

- There is a strong upward trend. This trend is non-stationary (i.e. not deterministic) according to the statistical test(s): Augmented Dickey-Fuller. The test(s) KPSS 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) 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 does not improve 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 8.22% on the test set. Including the trend variable leads to a score equal to 15.83%.

Long-term Growth

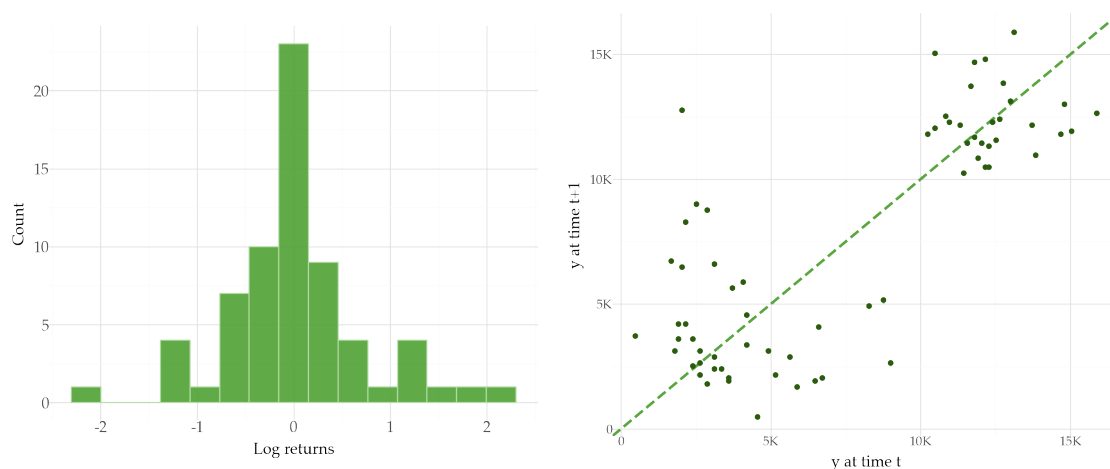


Figure 6: Distribution of log differences (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram show the distribution of these changes using log returns. The lag-plot depicts the randomness in the data. The time series shows greater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

- The time series has an average growth (log returns) of 0.02 (median equal to -0.01). The volatility of the returns in terms of standard deviation is 0.71. The skewness of the log differenced series is equal to 0.16, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness. The excess kurtosis of the log differenced series is equal to 1.71. This indicates a heavy tailed distribution.
- Concerning the symmetry of returns, 44.78% of the log differences are positive. The average of positive returns is 0.56, while the average of negative returns (53.73% of all returns) is -0.43. Overall, there are 46 return direction changes (69.7% of the data points)
- In the tails, 4.48% of returns fall beyond 2 standard deviations from the mean. The largest positive return is 2.05 on November 1990. Conversely, the largest decline is -2.25 (on October 1990).
- **Preliminary experiments:** Modeling the time series of first differences does not seem to improve forecasting accuracy. Experiments were conducted using a LightGBM algorithm and evaluated using SMAPE loss function. Using the original time series led to a 8.22% SMAPE. The scores using the differenced and log differenced time series are 10.5% and 15.01%, 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)

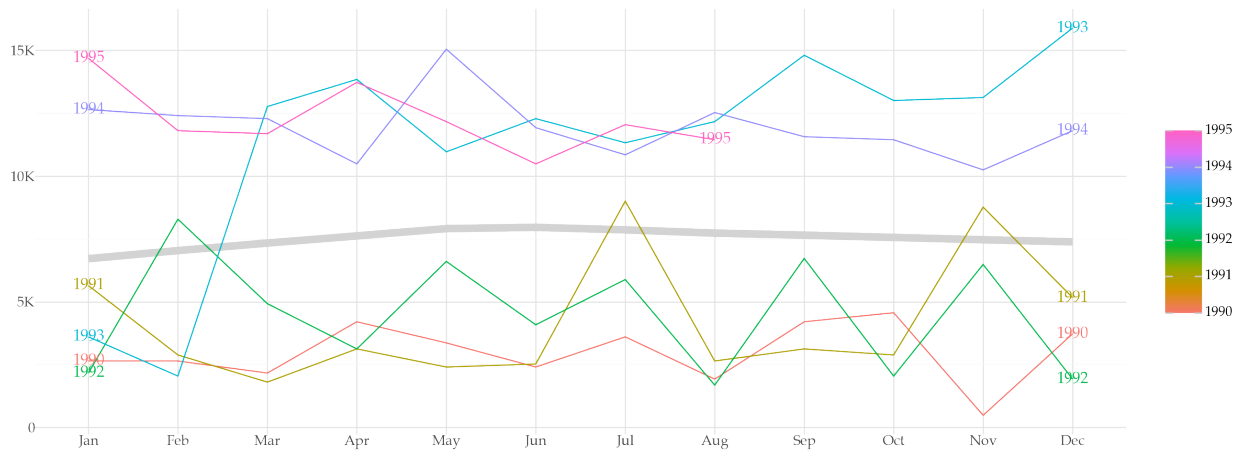


Figure 7: Seasonal plot of monthly values grouped by year.

- The seasonal strength is 0.27. 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 stationary in yearly seasonality.
- ****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 (Quarterly)

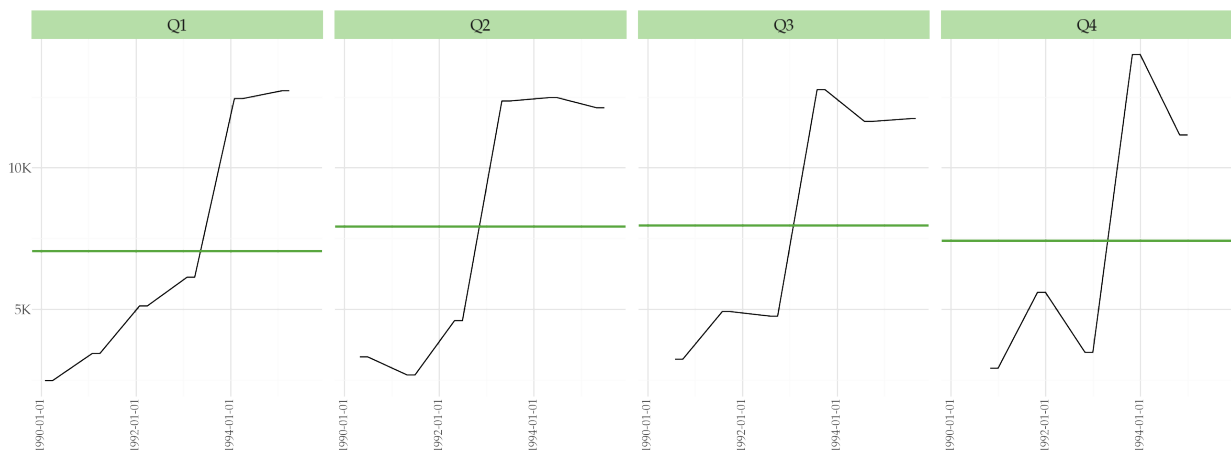


Figure 8: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- 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 strong 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.

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

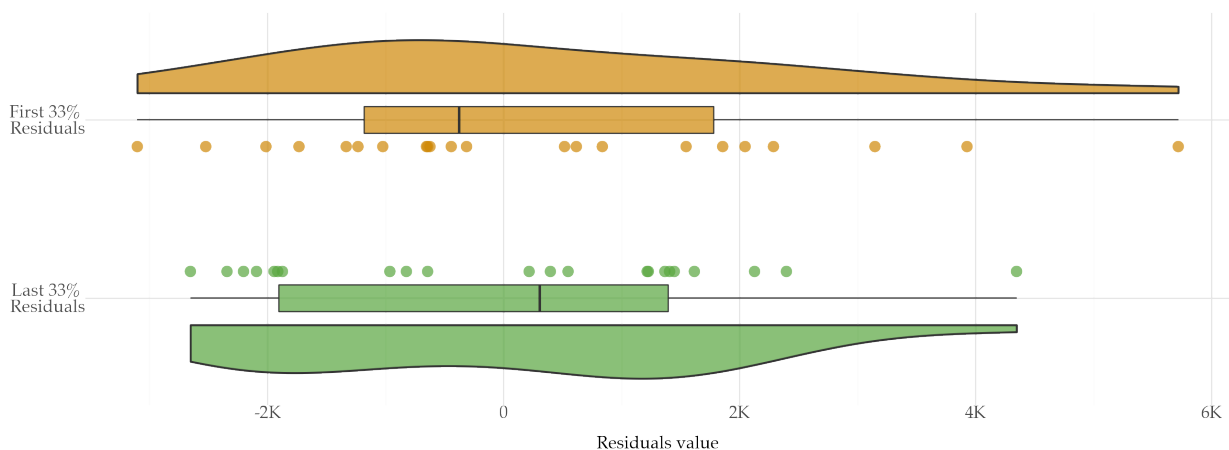


Figure 9: Time series residuals analysis based on a linear trend model. Difference in the distribution of the residuals in the first and last thirds of the series, following a Goldfeld-Quand partition.

- The following tests suggest that the time series is heteroskedastic: Breusch-Pagan. But, other tests (White 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 8.22% SMAPE using lag-based features):
 - Log returns: 15.01% SMAPE
 - Log transformation: 32.73% SMAPE
 - Box-Cox transformation: 20.42% 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

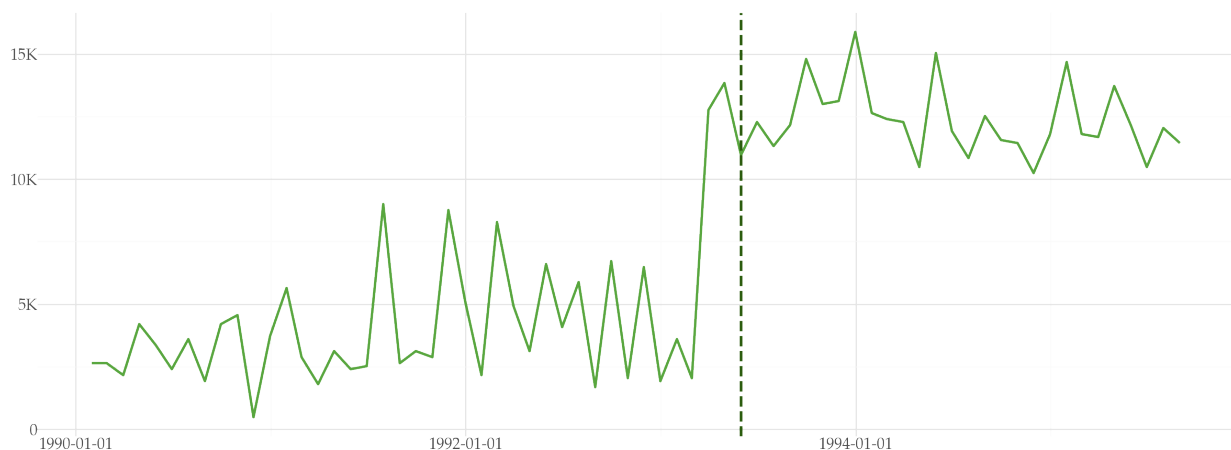


Figure 10: Time series plot with marked change points according to the PELT method.

- A single change point was found in the time series.
- The change point was found at May 1993 where the time series shows an decreasing level.

Effect on Model Parameters

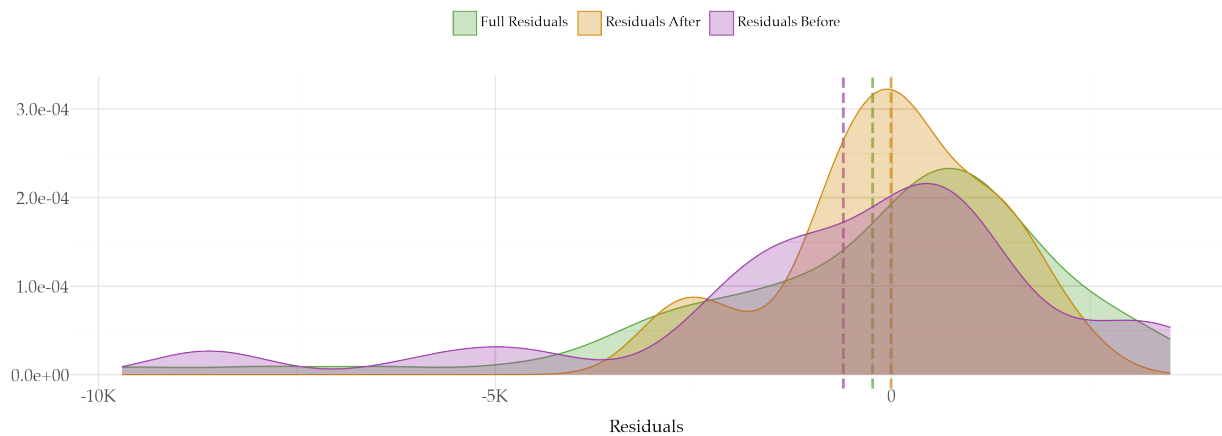


Figure 11: Distribution of the residuals of an ARIMA model before and after the first detected change point. The plot compares three kernel density estimates: residuals from the pre-change model, post-change model, and full series model. This comparison helps assess whether the structural break affects 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 8.22% remained the same when including the intervention (8.22%).