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: 2024-10-29 14:54

Series Name: M3

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

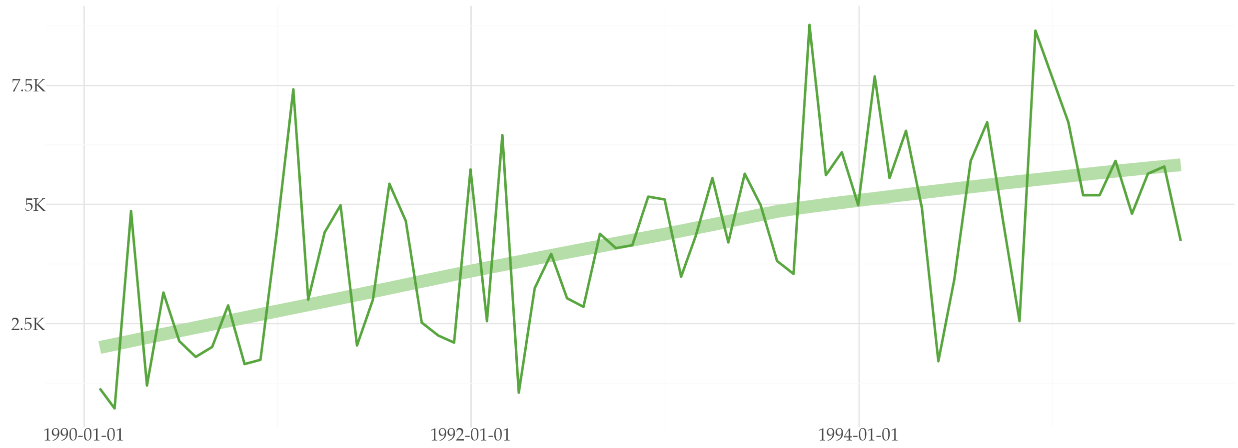

Figure 1: Time series line plot.

- A total of 68 monthly observations which span from January 1990 to August 1995.
- The mean value of the series is 4289.12 (median equal to 4395), with a standard deviation of 1882.65. The data ranges from a minimum of 720 to a maximum of 8760.

Data Distribution

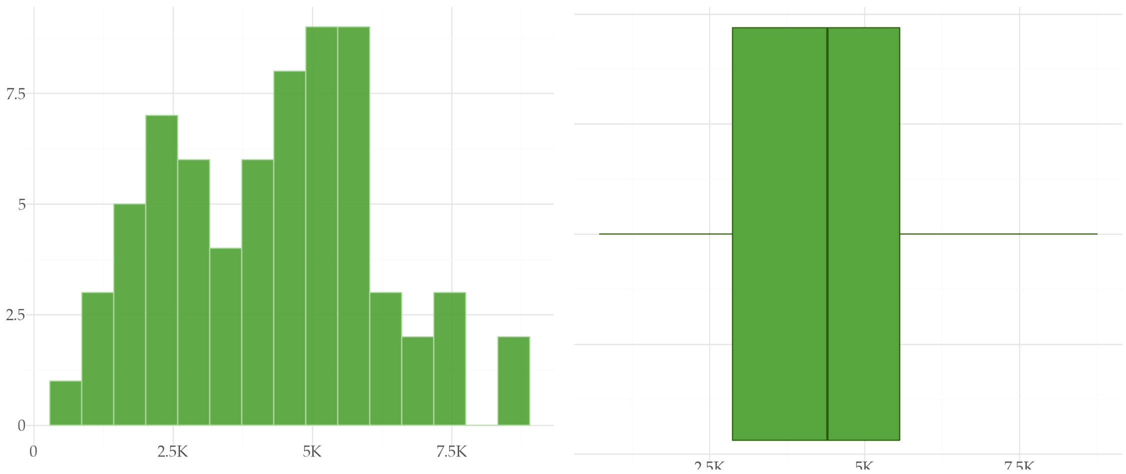

Figure 2: Distribution of the time series using an histogram (left) and a boxplot (right).

- The Kolmogorov-Smirnov test rejects the hypothesis that the series is distributed according to the following distributions: Power-law, Exponential, and Pareto
- The distribution with largest p-value is Gaussian (p-value equal to 0.96). 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, Logistic, Chi-squared, and Cauchy.

- There are no outliers in the data set according to the boxplot representation.
- The excess kurtosis is equal to -0.48. This value is similar to that found from data following a Gaussian distribution
- The skewness is equal to 0.17, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.

Trend and Seasonality

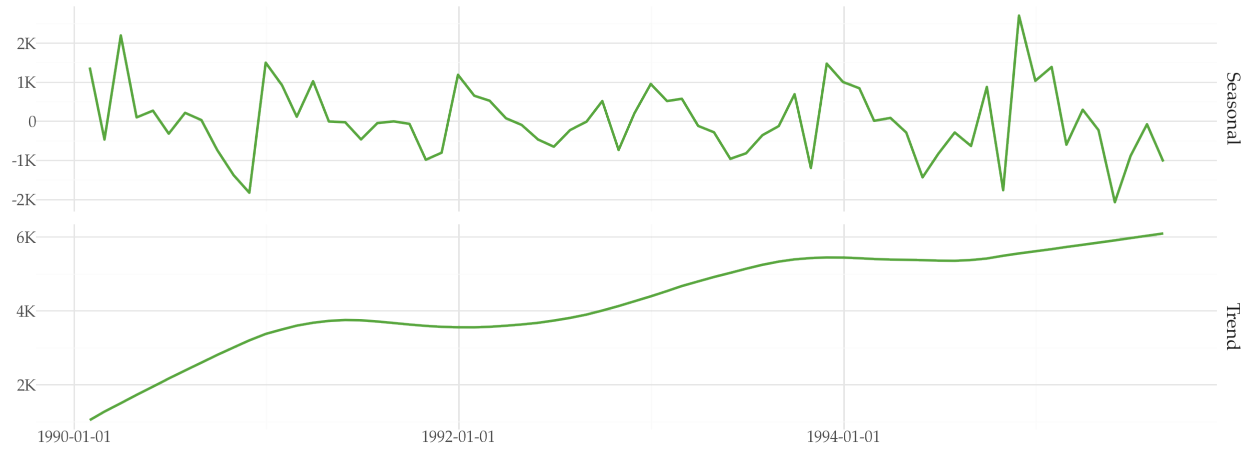

Figure 3: Seasonal and Trend components after decomposition using the STL (Season-Trend decomposition using LOESS) method.

- All hypothesis tests carried out (KPSS, Augmented Dickey-Fuller, and Philips-Perron) indicate that the time series is stationary in trend/level.
- 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 the hypothesis that the data is stationary

Auto-Correlation

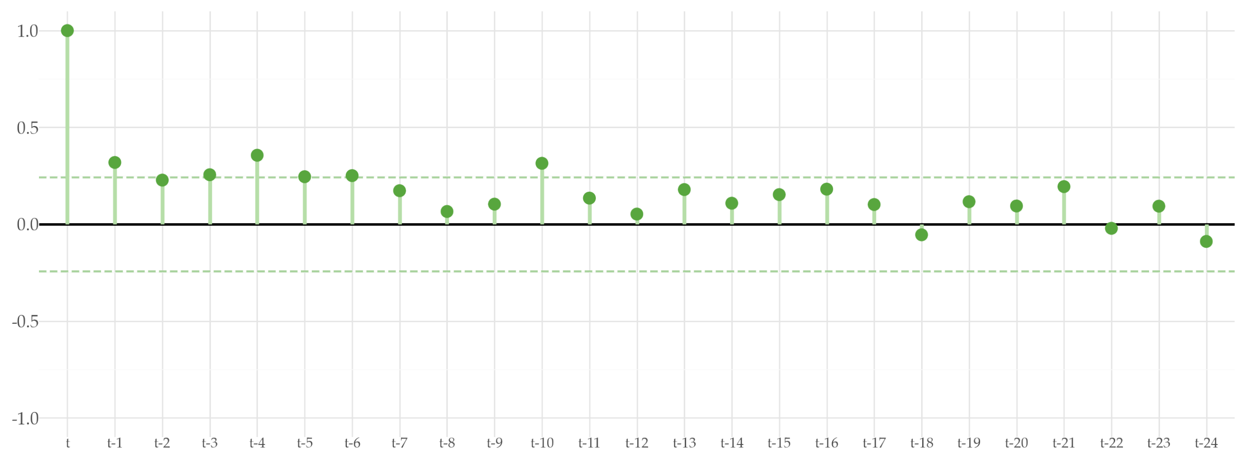

Figure 4: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-1, t-3, t-4, t-5, t-6, and t-10. 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

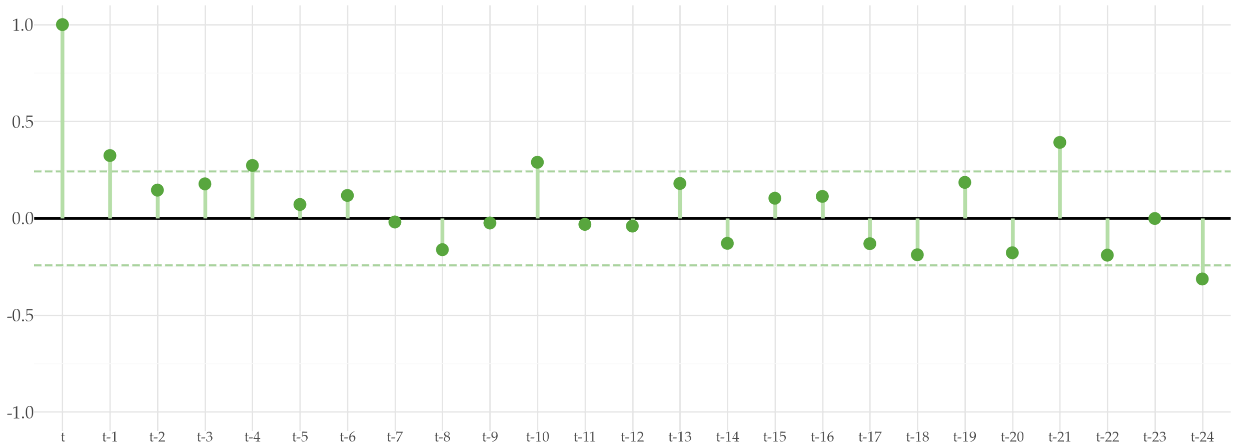

Figure 5: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-4, t-10, t-21, 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

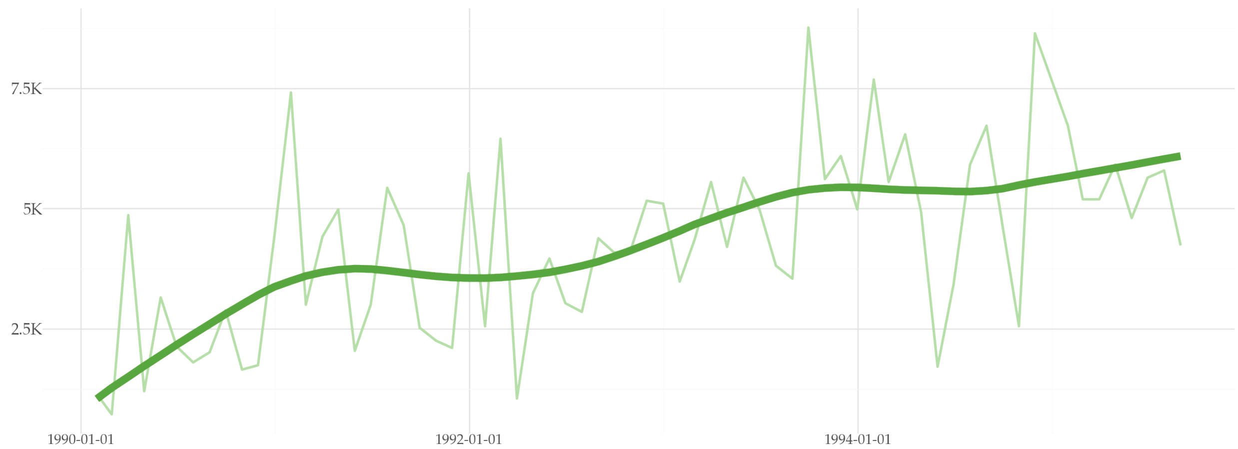

Figure 6: Time series trend plot.

- The time series has non-stationary trend according to the statistical test(s): . There is a moderate upward trend.
- On the other hand, the methods 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. All methods failed to reject that hypothesis.
- Including a trend explanatory variable which denotes the position of each observation does not improve forecasting performance.

Distribution of Differences

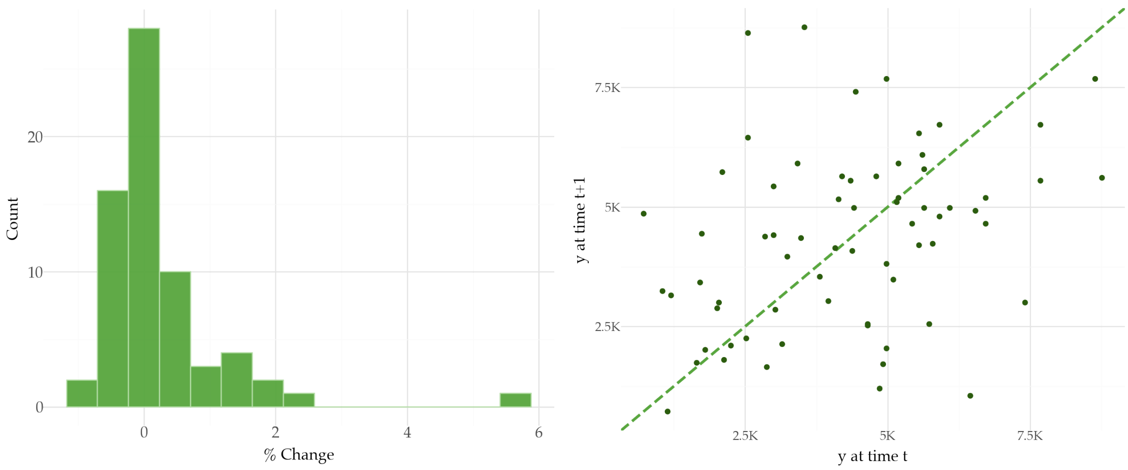

Figure 7: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The

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 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 excess kurtosis of the differenced series is equal to 0.42. This value is similar to that found from data following a Gaussian distribution
- The skewness of the differenced series is equal to 0.23, which is close to zero. This indicates a symmetric distribution, though there is a slight right skewness.
- **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)

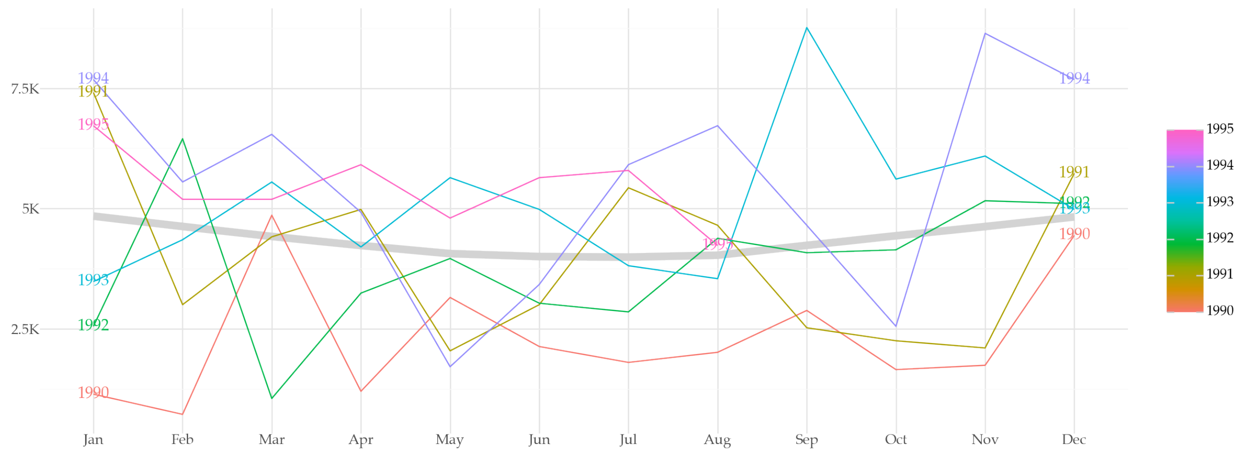

Figure 8: Seasonal plot of monthly values grouped by year.

- 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)

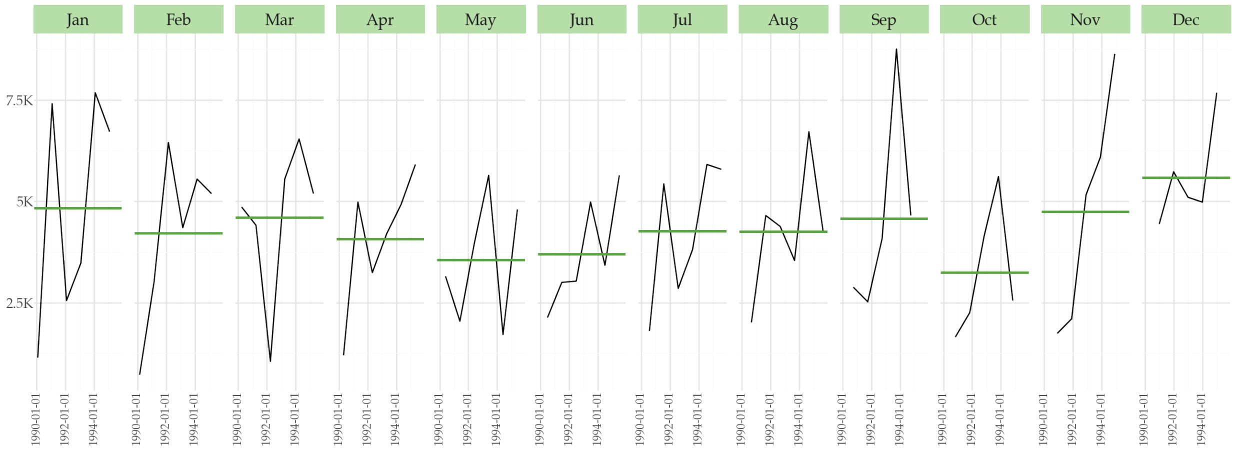

Figure 9: Monthly seasonal sub-series. This plot helps to understand how the data varies within and across monthly groups.

- 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)

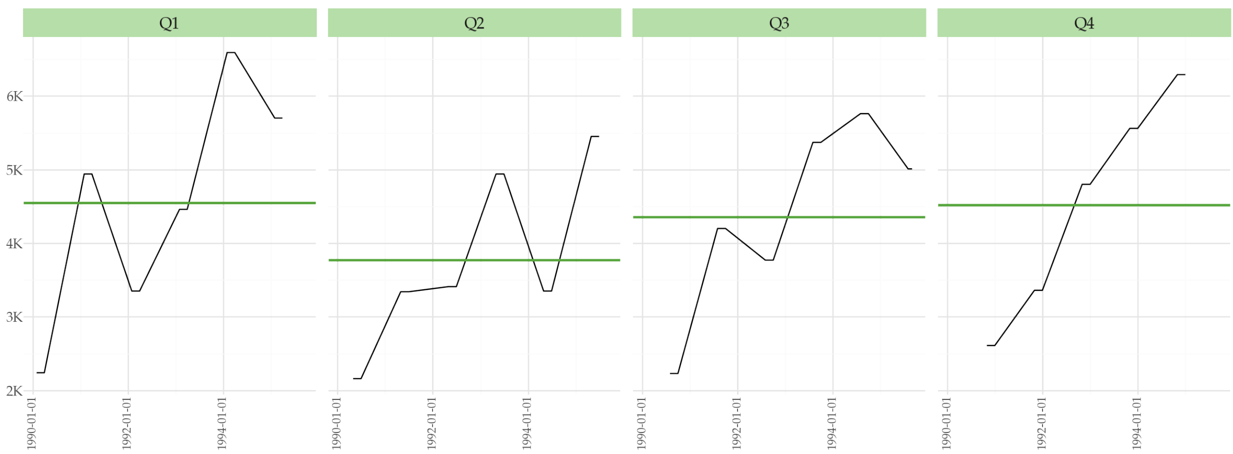

Figure 10: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- 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. Yet, including information about this period in the forecasting model decreased its performance.