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 23:32

Series Name: M5

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

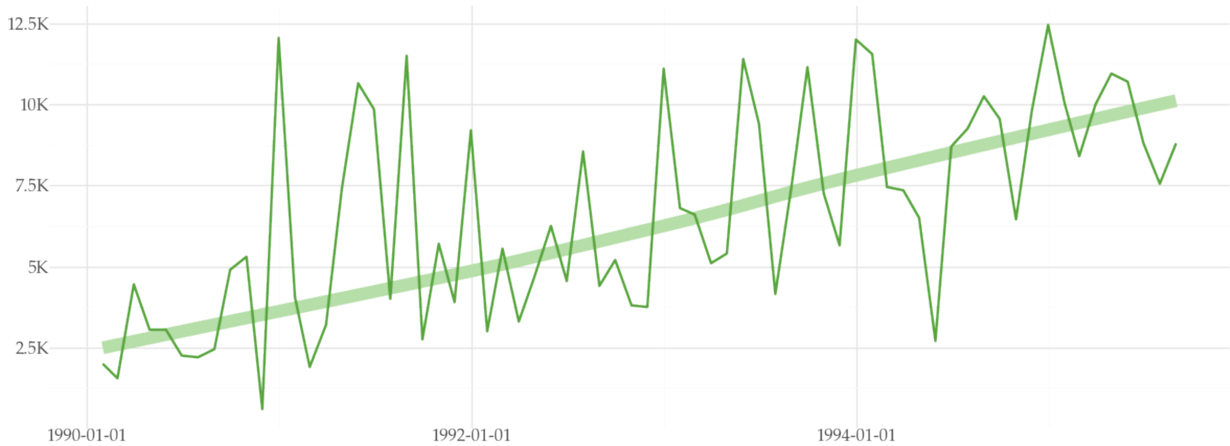

Figure 1: Time series line plot.

- A total of 68 observations spanning from January 1990 to August 1995. These are collected with a monthly sampling frequency.
- The data ranges from a minimum of 600 to a maximum of 12450, starting in 2000 and ending in 8800 during the observed period. The average growth percentage per observation is 44.51% (median equal to -2.22%), with an average value of 6643.38. There are no missing values in the time series.

Trend, Seasonality, and Residuals

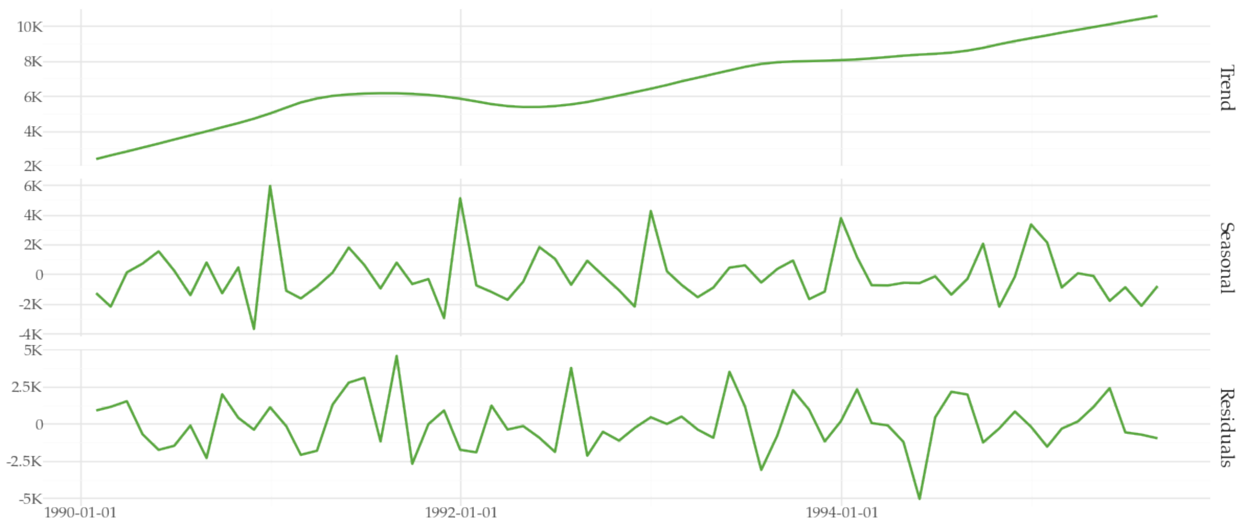

Figure 2: Seasonal, Trend, and Residuals components after decomposition on a monthly frequency using the STL (Season-Trend decomposition using LOESS) method.

- The trend strength is 0.59 (ranges from 0 to 1). All hypothesis tests carried out (KPSS, Augmented Dickey-Fuller, and Philips-Perron) indicate that the time series is stationary in trend or level.

- The seasonal strength is 0.52 (ranges from 0 to 1). Tests for yearly seasonal stationarity show mixed results: the OCSB test indicates presence of a seasonal unit root, while the Wang-Smith-Hyndman test suggests stationarity.

- The STL decomposition residuals show unbalanced behavior: 42.65% of residuals are positive and 57.35% negative. The average magnitude of positive residuals is 1554.335 compared to -1144.419 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

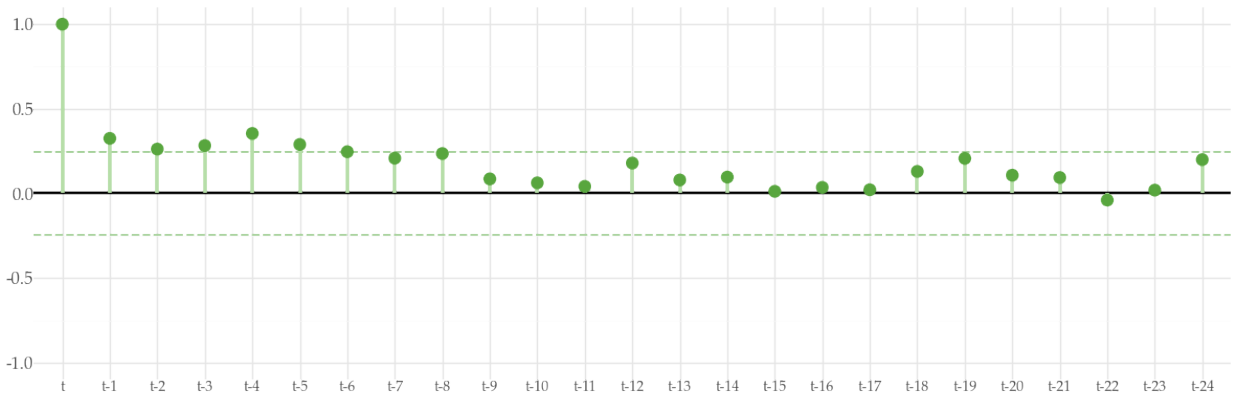

Figure 3: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-1, t-2, t-3, t-4, t-5, and t-6. The autocorrelation is positive for all lags with a significant value.

- The following seasonal lags show a significant autocorrelation: t-24

Partial Auto-Correlation

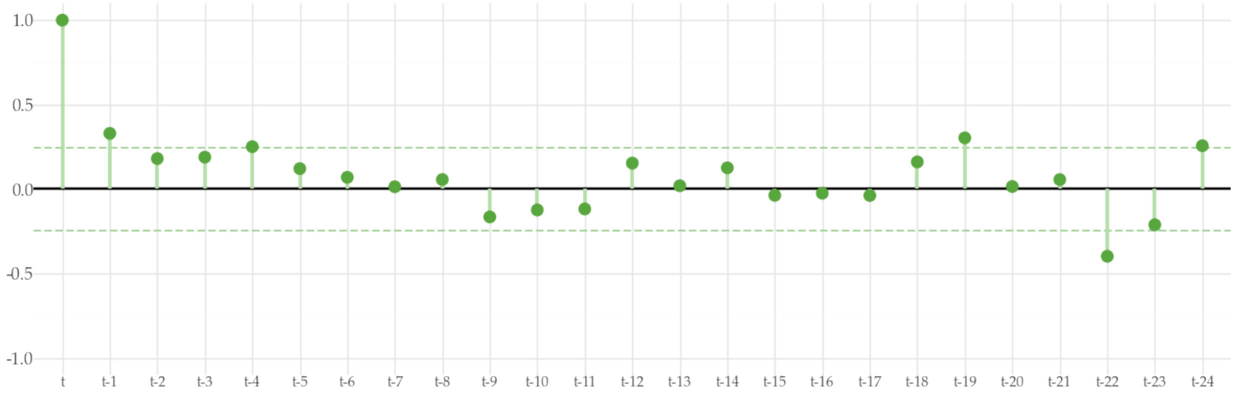

Figure 4: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-4, t-19, t-22, and t-24.
- The following seasonal lags show a significant partial autocorrelation: t-24

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

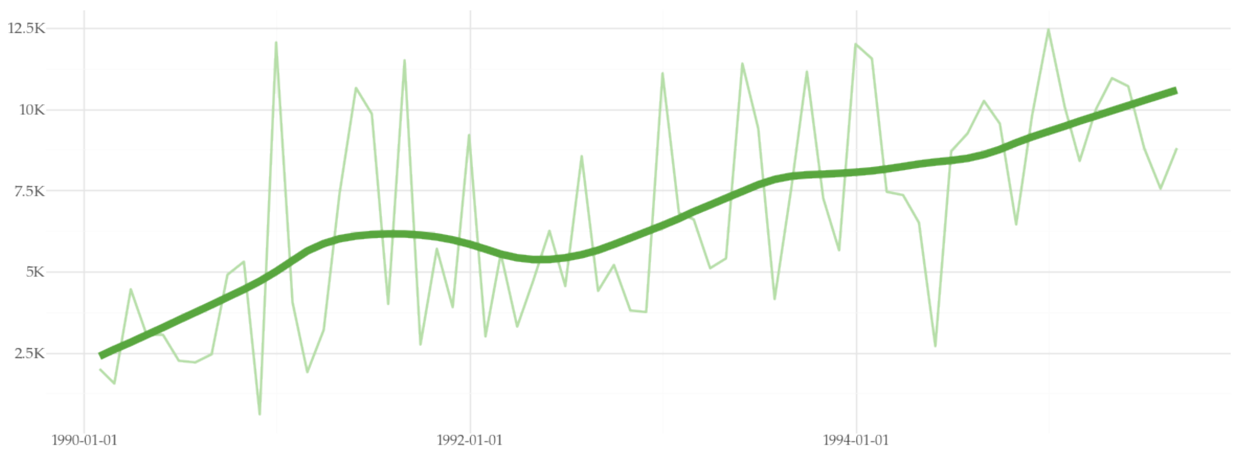

Figure 5: Time series trend plot.

- There is a moderate upward trend. The tests KPSS, Augmented Dickey-Fuller, 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. The following test reject this hypothesis: KPSS. But, the tests Augmented Dickey-Fuller and Philips-Perron suggest 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 37.96% on the test set. Including the trend variable improved the SMAPE to 28.01%.

Long-term Growth

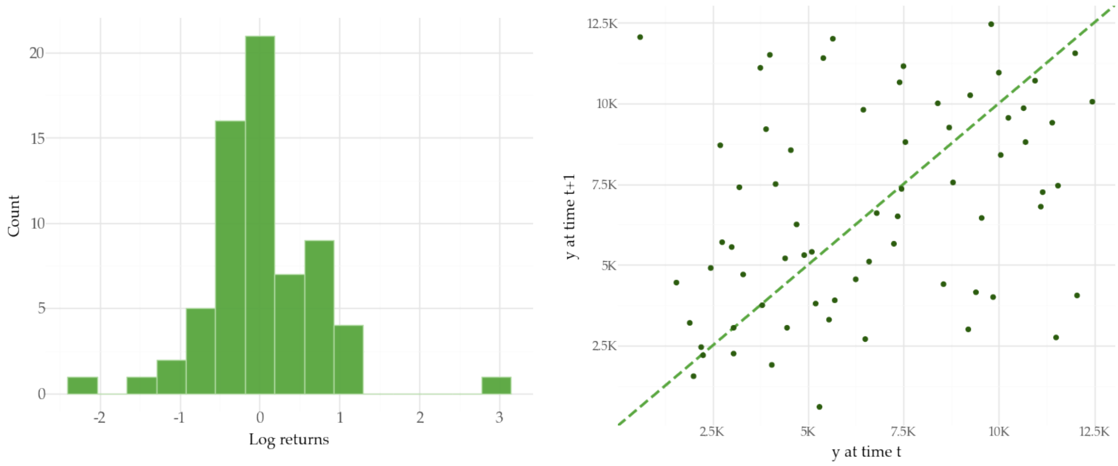

Figure 6: Distribution of log differences (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram show the distribution of these changes using log returns. The lag-plot depicts the randomness in the data. The time series shows greater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

- The time series has an average growth (log returns) of 0.02 (median equal to -0.02). The volatility of the returns in terms of standard deviation is 0.72. The skewness of the log differenced series is equal to 0.59, indicates that the right tail is long relative to the left tail. The excess kurtosis of the log differenced series is equal to 3.93. 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 0.58, while the average of negative returns (53.73% of all returns) is -0.44. Overall, there are 37 return direction changes (56.06% of the data points)

- In the tails, 4.48% of returns fall beyond 2 standard deviations from the mean. The largest positive return is 3.0 on November 1990. Conversely, the largest decline is -2.18 (on October 1990).

- **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 37.96% SMAPE. The scores using the differenced and log differenced time series are 30.79% and 29.08%, 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)

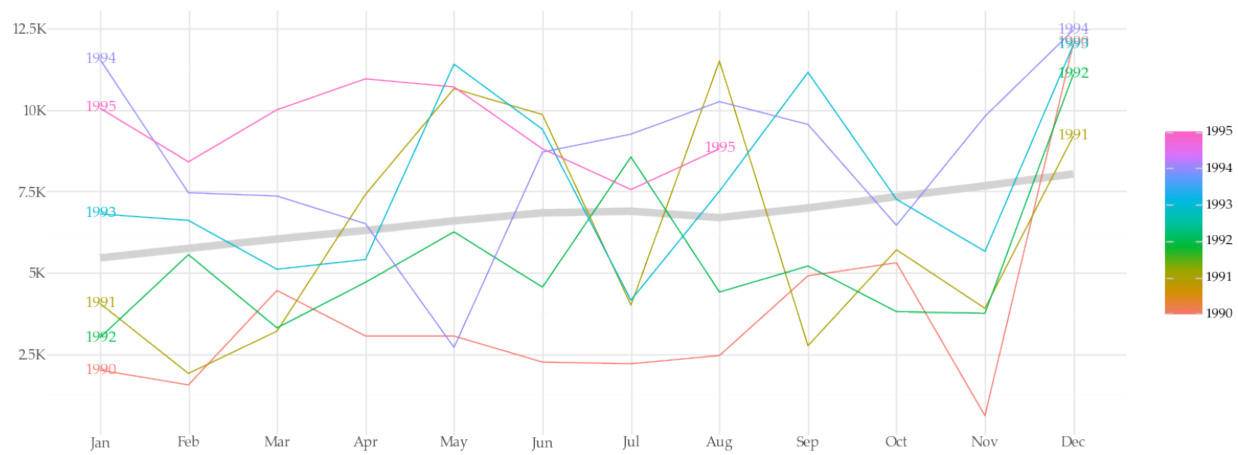

Figure 7: Seasonal plot of monthly values grouped by year.

- The seasonal strength is 0.52. This score ranges from 0 to 1 and values above 0.64 are considered significant. 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.
- **Preliminary experiments:** Modeling yearly patterns can improve forecast accuracy. Different approaches were tested relative to a base model using only lag-based features (37.96% SMAPE):
    - Fourier terms: 23.48% SMAPE
    - Seasonal differencing: 27.61% SMAPE
    - Monthly time features: 37.96% SMAPE

Seasonal Sub-series Plot (Quarterly)

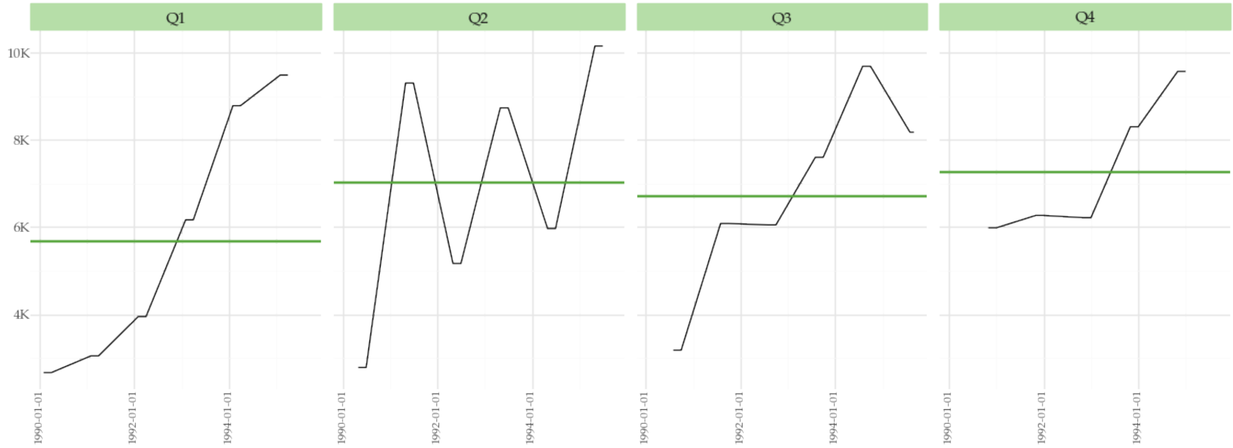

Figure 8: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- Statistical analysis of quarterly data shows no evidence of systematic differences across quarters, with both means (Kruskal-Wallis test) and variances (Levene's test) being statistically similar.

- 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 (37.96% SMAPE):
    - Fourier terms: 33.56% SMAPE
    - Quarterly seasonal differencing: 38.32% SMAPE
    - Quarterly time features: 37.96% SMAPE