Basic Attributes

Time Series Plot

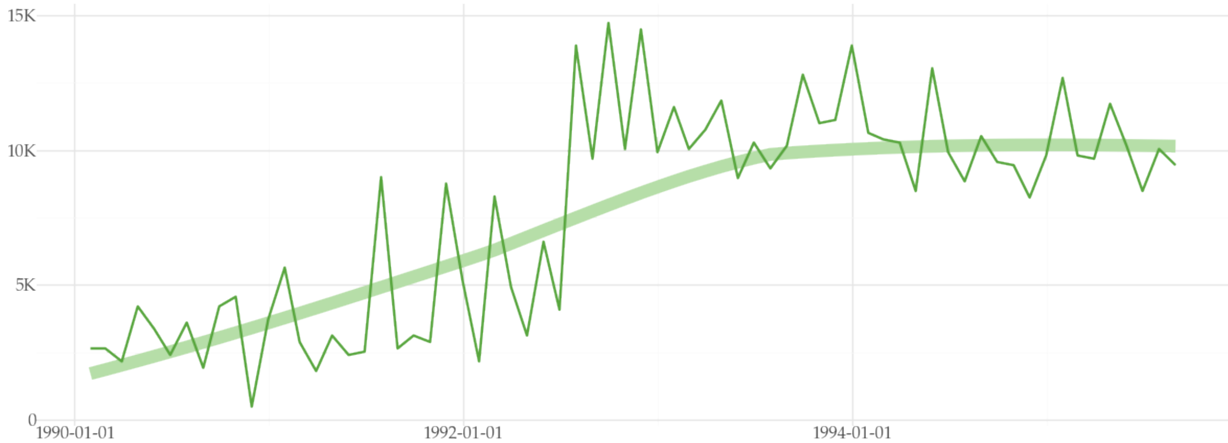

Figure 1: Time series line plot.

- A total of 68 monthly observations which span from January 1990 to August 1995.
- The mean value of the series is 7655.88 (median equal to 8980), with a standard deviation of 3887.89. The data ranges from a minimum of 480 to a maximum of 14720.

Data Distribution

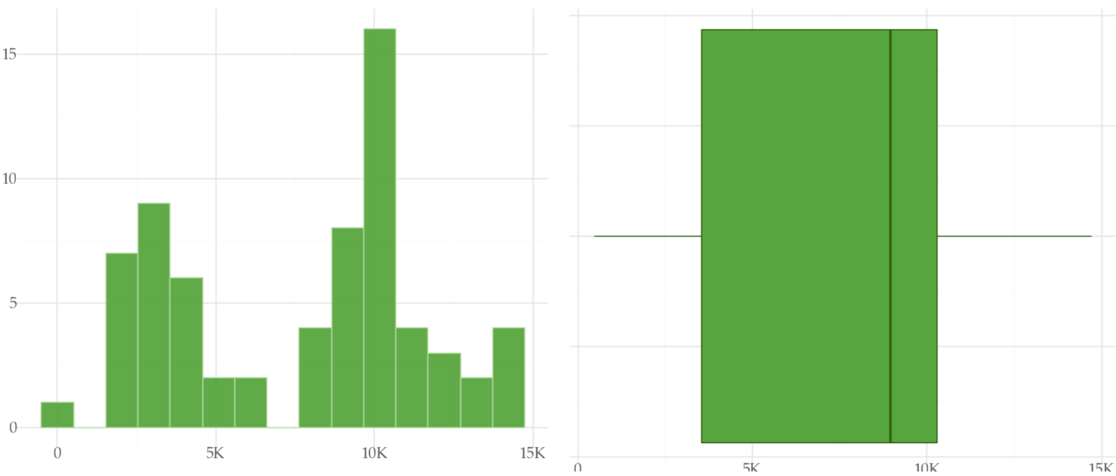

Figure 2: Distribution of the time series using an histogram (left) and a boxplot (right).

- The Kolmogorov-Smirnov test rejects the hypothesis that the series is distributed according to the following distributions: Gaussian, Log-Normal, Exponentially Modified Gaussian distribution, Chi-squared, Gamma, Cauchy, Power-law, Exponential, and Pareto
- The distribution with largest p-value is Logistic (p-value equal to 0.1). No other distribution was found with an acceptable significance.
- There are no outliers in the data set according to the boxplot representation.
- The excess kurtosis is equal to -1.25. This indicates that the data has a light tailed distribution.
- The skewness is equal to -0.17, which is close to zero. This indicates a symmetric distribution, though there is a slight left skewness.

Trend and Seasonality

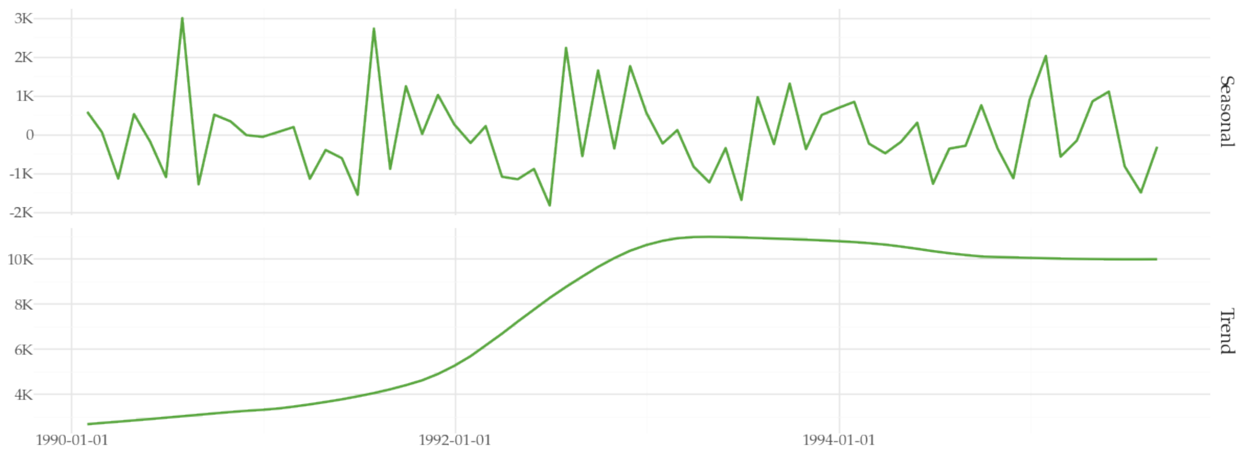

Figure 3: Seasonal and Trend components after decomposition using the STL (Season-Trend decomposition using LOESS) method.

- The following tests indicate that the time series is non-stationary in trend/level: KPSS and Augmented Dickey-Fuller. On the other hand, other tests (PP) fail to reject the hypothesis that the data is stationary
- The following tests indicate that the time series is non-stationary in seasonality for the specified period: HEGY and Canova-Hansen. On the other hand, other tests (Wang-Smith-Hyndman and OCSB) fail to reject the hypothesis that the data is stationary

Auto-Correlation

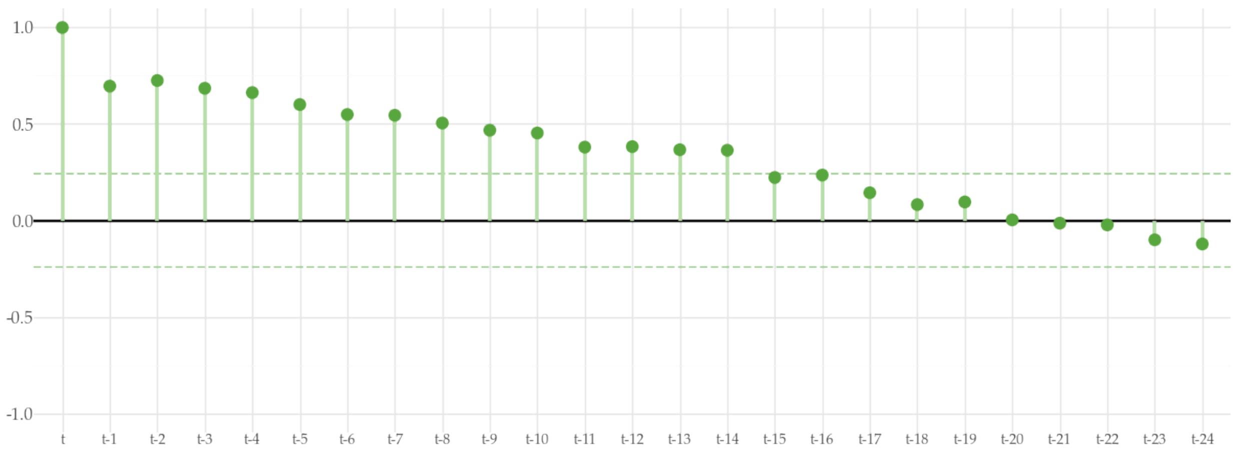

Figure 4: Auto-correlation plot up to 24 lags.

- The following lags show significant autocorrelation: t-1, t-2, t-3, t-4, t-5, t-6, t-7, t-8, t-9, t-10, t-11, t-12, t-13, and t-14. The autocorrelation is positive for all lags with a significant value.
- Only the following lags relative to the seasonal period show a significant autocorrelation: t-12

Partial Auto-Correlation

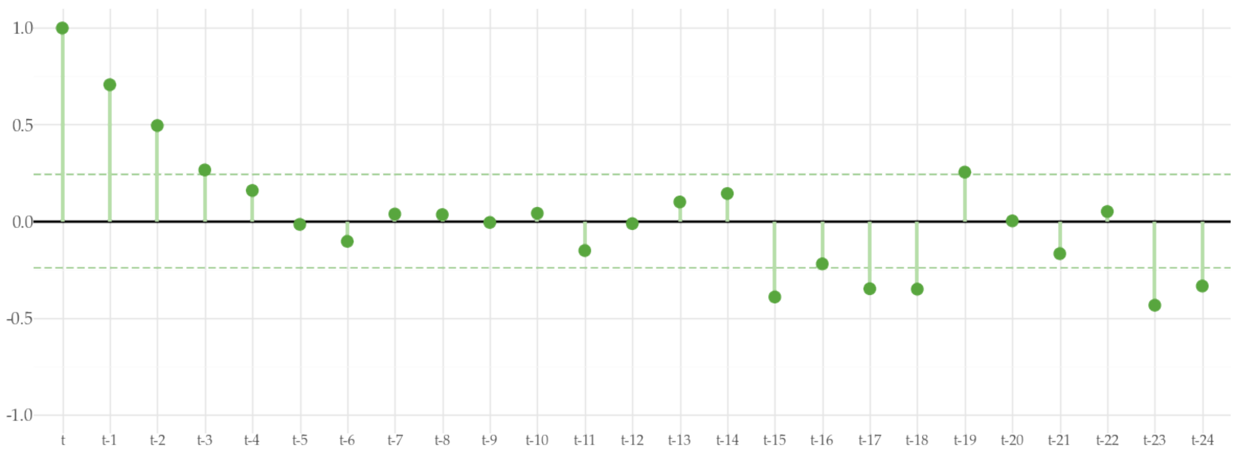

Figure 5: Partial Auto-correlation plot up to 24 lags. At each lag, the partial auto-correlation takes into account the previous correlations.

- The following lags show significant partial autocorrelation: t-1, t-2, t-3, t-15, t-17, t-18, t-19, t-23, and t-24.
- None of the lags relative to the seasonal period (t-12 and t-24) show any significant partial autocorrelation.

Report Organization

The following sections dive deeper into some relevant aspects of the time series. The remaining sections address the following topics. Some analysis are based on forecasting experiments. These were carried out using a ridge regression model trained for one-step ahead forecasting.

- Section 2 details the analysis of trend. This analysis is split into two parts. The statistical analysis that described whether the time series is trend-stationary or not are described. Besides, we test whether some typical transformation used to deal with trend lead to better forecasting performance.
- The seasonal component is analised in Section 3. Several statistical tests are carried out for different seasonal periods. These evaluate not only seasonal stationarity but also statistical differences among seasonal groups. Finally, tentative experimental results are reported which indicate possible directions for modeling seasonality.
- Section 4 presents the analysis of the variance. Different heteroskedasticity tests are carried out. Besides these, experiments are performed to assess whether transforming the data improves forecasting performance.
- Finally, change detection results are described in Section 5. The distribution of the data is analysed before and after change occurs.

Trend

The trend component refers to the long-term change in the mean level of the series.

Trend Line Plot

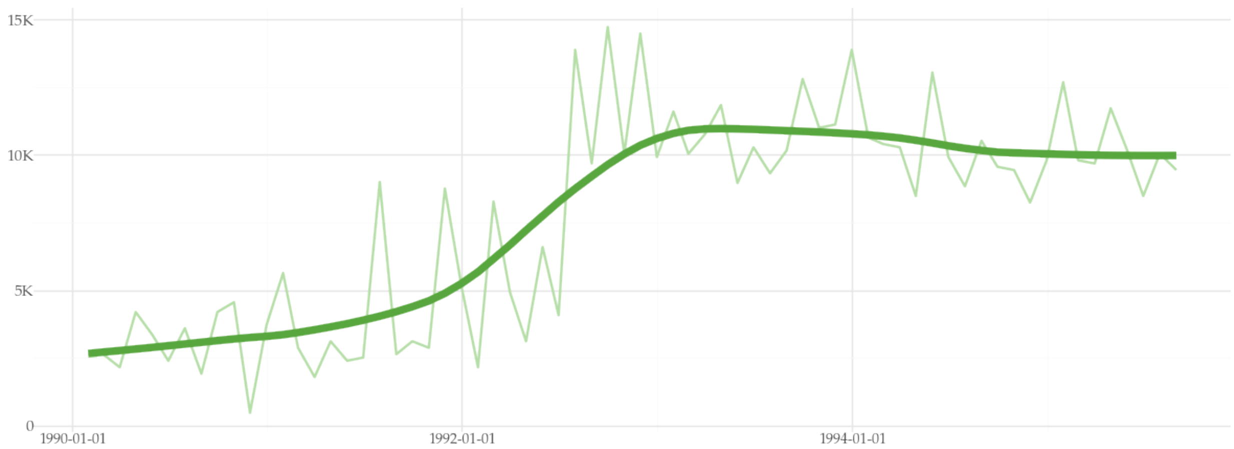

Figure 6: Time series trend plot.

- The time series has non-stationary trend according to the statistical test(s): KPSS and Augmented Dickey-Fuller. There is a strong upward trend.
- On the other hand, the method PP 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 level. The method(s) KPSS and Augmented Dickey-Fuller reject this hypothesis. But, the test(s) PP fail to reject.
- Including a trend explanatory variable which denotes the position of each observation improves forecasting performance.

Distribution of Differences

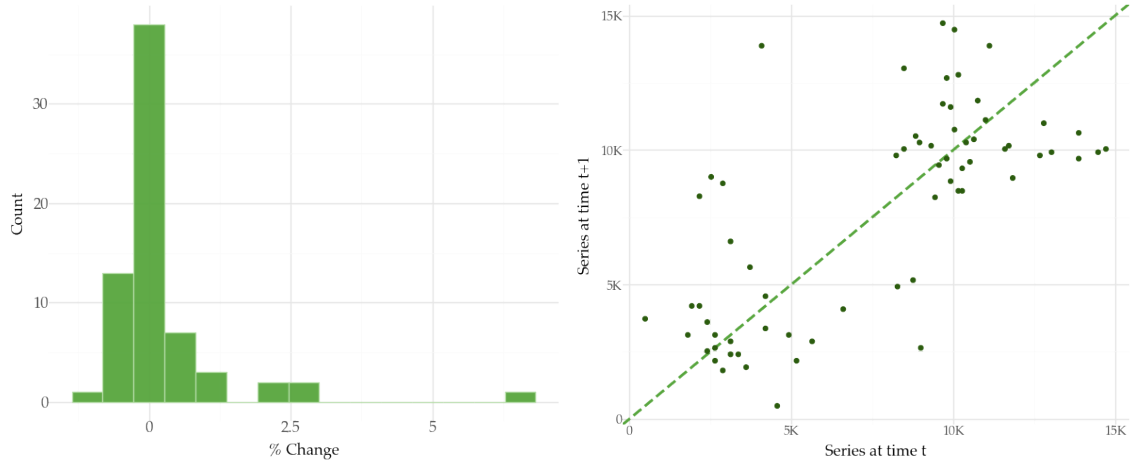

Figure 7: Distribution of percentage changes (left), and a Lag-plot (right). These plots help to understand how the data changes over consecutive observations. The histogram shown the distribution of these changes. The lag-plot depicts the randomness in the data. The time series shows greater randomness 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 0.98). 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 0.73. This value is similar to that found from data following a Gaussian distribution
- The skewness of the differenced series is equal to 0.62, indicates that the right tail is long relative to the left tail.
- **Forecasting experiments**: Taking first differences improves forecasting performance.

Seasonality

Seasonality represents regular and predictable patterns which recur in a fixed period (e.g. every month).

Seasonal Line Plot (Monthly)

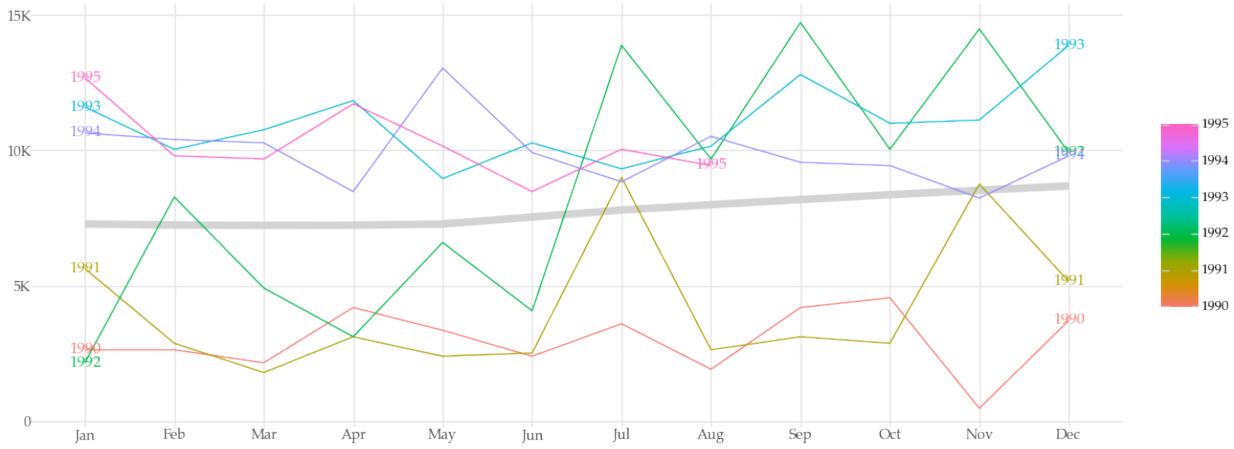

Figure 8: Seasonal plot of monthly values grouped by year.

- The following tests indicate that the time series is non-stationary in seasonality for a yearly period: HEGY and Canova-Hansen. On the other hand, other tests (Wang-Smith-Hyndman and OCSB) 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)

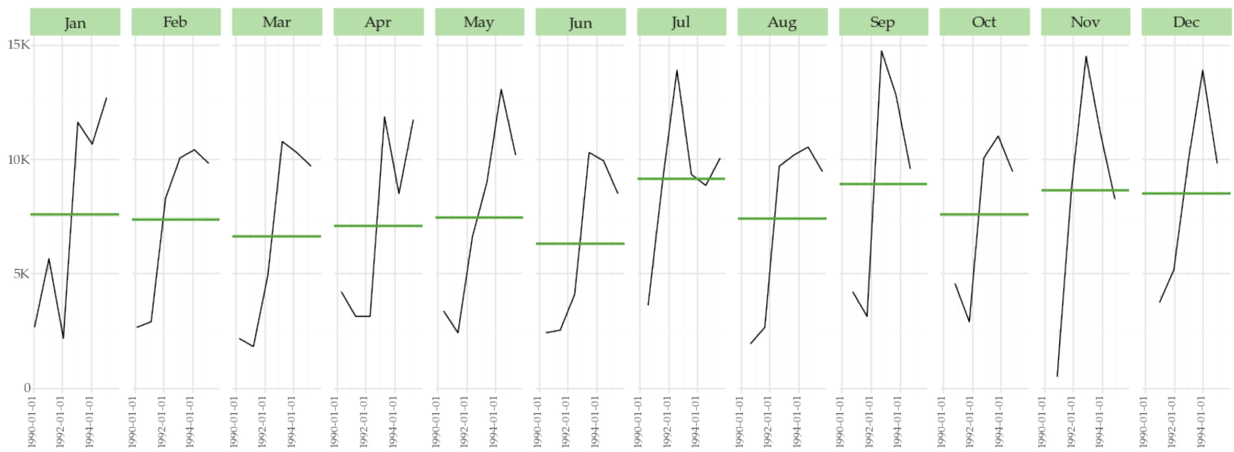

Figure 9: Monthly seasonal sub-series. This plot helps to understand how the data varies within and across monthly groups.

- 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. Within each group, there is also indication that the data is not constant around a level in 66% of the Months.

- **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)

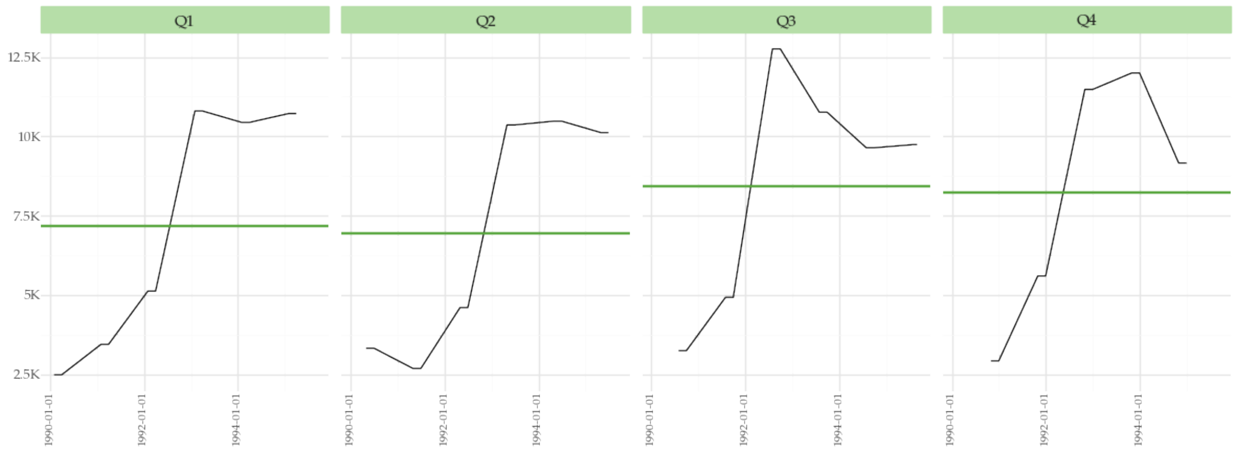

Figure 10: Quarterly seasonal sub-series. This plot helps to understand how the data varies within and across quarterly groups.

- 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: HEGY and Canova-Hansen. On the other hand, other tests (Wang-Smith-Hyndman and OCSB) fail to reject the stationary null hypothesis.
- Overall, there is a reasonable evidence that the time series is not stationary around a constant level. Within each group, there is also indication that the data is not constant around a level in all of the Quarters.
- **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 quantifies the dispersion of the time series. The data set is heteroskedastic if the variance changes over time, or homoskedastic otherwise.

Heteroskedasticity Testing

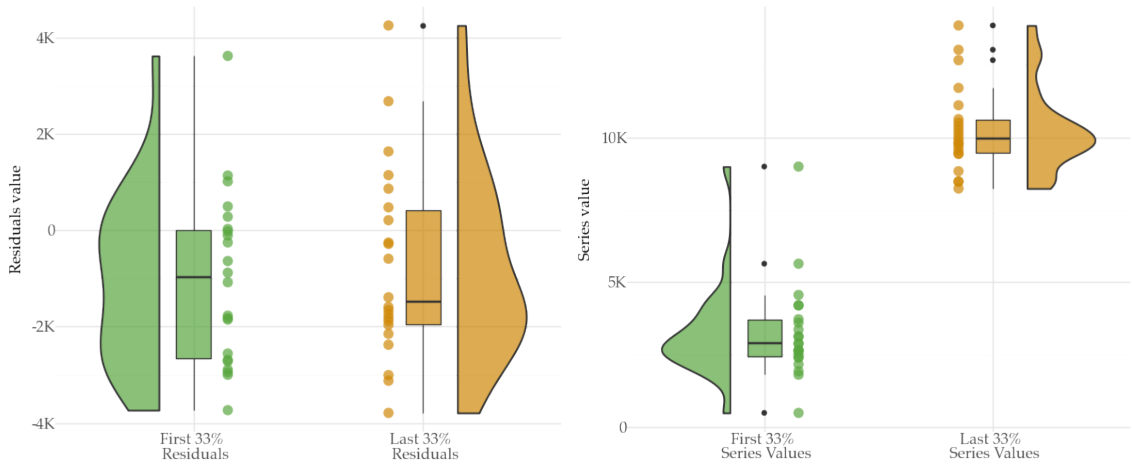

Figure 11: Time series residuals analysis. Difference in the distribution of the residuals (left) and the series (right) in the first and last thirds of the series, following a Goldfeld-Quand partition.

- In the analysis of seasonality we did not find significant differences in the dispersion among periodic groups of observations.
- The following tests suggest that the time series is heteroskedastic: Breusch-Pagan and Goldfeld-Quandt. But, other tests (White) fail to reject the hypothesis that the time series has a constant variance.
- **Forecasting experiments**: Transforming the series with either the logarithm or the Box-Cox method did not improve forecasting performance.

Change Detection

Change points occur when the time series distribution changes. These points complicate the application a consistent methodology for time series analysis.

Change Points

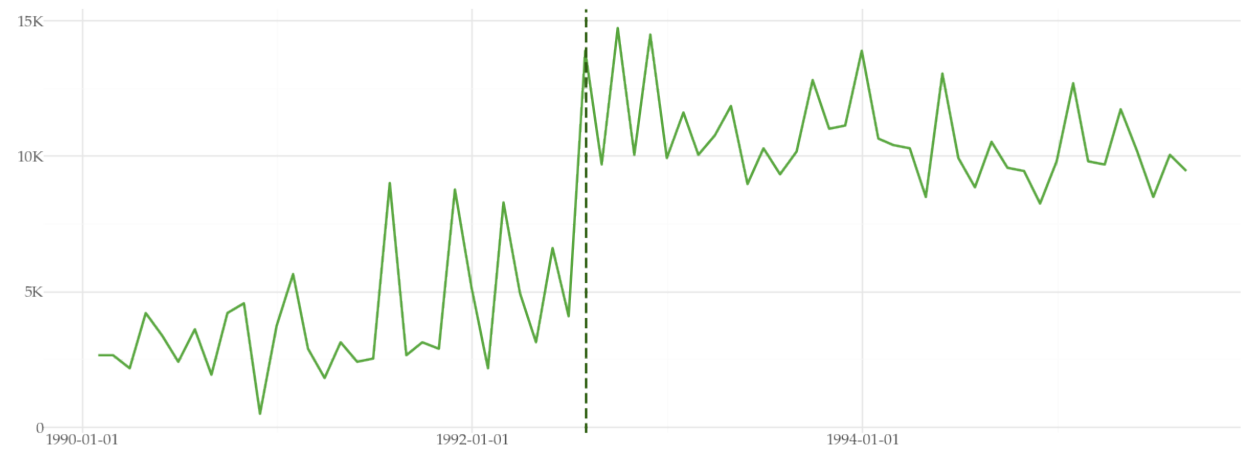

Figure 12: Time series plot with marked change points according to the PELT method.

- A single change point was found in the time series.
- The change point was found at July 1992 where the time series shows an decreasing tendency.

Changes in Distribution

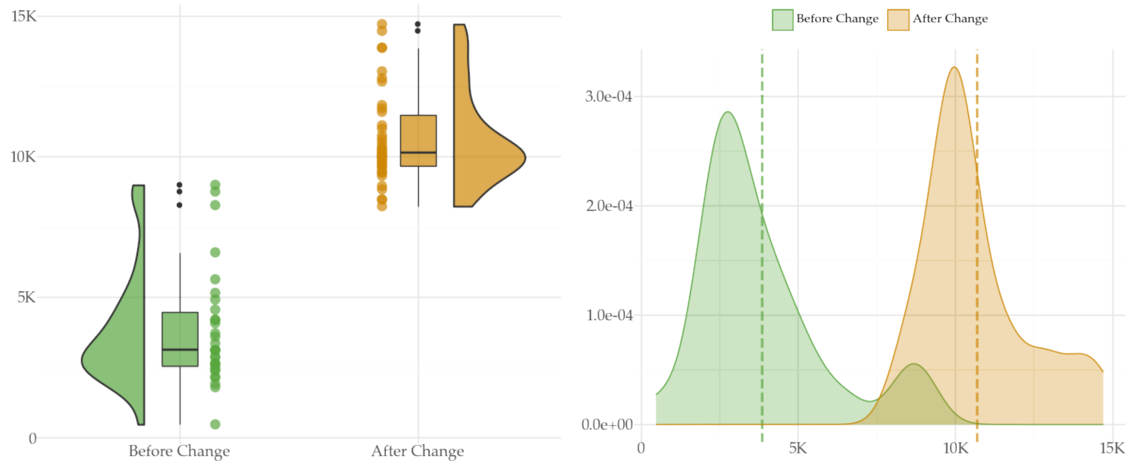

Figure 13: Time series analysis before and after the first detected change point occurs. Paired distributions before and after change (left), and overlapped density plot (right).

- The distribution before and after the first change point are significantly different.