

1. Problem

Which of the following statements IS correct?

- (a) The structural break can be detected by comparing heterogeneity of the the time series subsets with heterogeneity of the entire series
- (b) Structural break detection algorithms should be applied to initial time series
- (c) Usually it's enough to consider only one structural break
- (d) There can only be one correct to identify the structural break

Solution

- (a) True.
- (b) False. They can be applied to transformed data as well as to individual components
- (c) False. This depends on data, multiple structural breaks can be as probable as single ones.
- (d) False. Point of changes of the series behavior is subjective and depends on data and underlying model.

2. Problem

Consider a sample $y^T = [4.22, 666, 4.09, 4.12, 1.72, 5.44]$. Say, you want to detect an outlier based on the residuals of a model on a constant $y_t = c + u_t$, where u_t is a white noise. Give a new value for an outlier rounded to 2 decimal digits as an answer if there is one, else give 0 as an answer.

Solution

Check Anomaly detection algorithm from lecture "Anomaly detection".

3. Problem

Consider tsAirgap dataset on monthly totals of international airline passengers with missing values filled in with linear interpolation.

Let's say 37th observation of that series is equal to 0.

Remove anomalies based on STL decomposition. Provide the absolute value of mean difference of the initial 37th observation and the imputed value rounded with 3 decimal digits as an answer.

You can load the tsAirgap data set in R by issuing the following command at the console `data("tsAirgap")` after importing "imputeTS" library or if you use other programming languages you can download it [here](#).

Solution

You can use `time_decompose()` `%>% anomalize(remainder) %>% time_recompose()` and `clean_anomalies()` functions.

4. Problem

Consider tsAirgap dataset on monthly totals of international airline passengers with missing values. You need to fill in the missing data using (1) Kalman Filtering based on ARIMA model with automatically chosen parameters and (2) linear interpolation. Provide the absolute value of mean difference of the imputed values rounded with 3 decimal digits as an answer.

You can load the tsAirgap data set in R by issuing the following command at the console `data("tsAirgap")` after importing "imputeTS" library or if you use other programming languages you can download it [here](#).

Solution

You can use `na_kalman(ts, model = 'auto.arima')` and `na_interpolation(ts)` function.

5. Problem

Select ALL methods that can be used to fill in the missing data:

- (a) Linear interpolation of the seasonal component of the STL decomposition
- (b) Filling in with the conditional mathematical expectation using Kalman filter
- (c) Random number drawn from Cauchy distribution
- (d) Fill in missing values with previous value

Solution

- (a) False. Linear interpolation of the deseasonalized component of the STL decomposition + a seasonal component
- (b) True.
- (c) False.
- (d) True.

6. Problem

Consider a sample $y^T = [1.22, 1.61, 2.69, NA, 2.53, 4.97]$ with missing value. Fill in the missing value using linear interpolation.

Provide the answer with 2 decimal digits.

Solution

Check lecture “Handling Missing Data”.

7. Problem

Which of the following statements ARE correct?

- (a) You should always remove the anomalous observations
- (b) Filling in the missing values with linear interpolation always give worse results than using a modelling approach
- (c) Structural breaks may relate to any sudden change in time series pattern, i.e. sudden volatility change
- (d) Whether an observation would be considered anomalous or not depends on an underlying model

Solution

- (a) False. Some models are robust to outliers, also categorizing the observations into anomalous and ordinary one is subjective in the first place
- (b) False. This depends on data
- (c) True.
- (d) True.

8. Problem

Consider tsAirgap dataset on monthly totals of international airline passengers with missing values filled in with linear interpolation.

Apply the log transform and then remove the linear trend. Check if there are any structural breaks using Binary Segmentation method. Give the number of structural breaks as an answer.

You can load the tsAirgap data set in R by issuing the following command at the console `data("tsAirgap")` after importing “imputeTS” library or if you use other programming languages you can download it [here](#).

Solution

You can use `cpt.mean(ts, method = 'BinSeg', Q = 1)` function.

Final Project

You are supposed to work with daily data on Russian spot electricity market. Choose one of the four time series from the dataset you are interested to work with.

1. (15 points) Visualize the series itself, the series of ordinary and seasonal differences, the components of the series, the ordinary and partial autocorrelation functions. Comment on the created plots.

2. (15 points) Prepare the data: fill in the missing values, remove outliers and correct for structural breaks if needed. Comment on the choice of the methods and the parameters.

3. (5 points) Is the series stationary? Support your answer with a suitable test.

4. (5 points) If it is reasonable to apply any transformation to the original series? Motivate your choice.

5. (50 points) Divide the series into test and train sets. Evaluate a set of models/algorithms on a test set. As a minimum, you should consider: SARIMA, ETS, OLS + a comparator of your choice. You are welcome to consider additional models as well. Provide a comment on the choice of model parameters!

6. (5 points) Choose the best model.

7. (5 points) Plot forecasts two years ahead by reevaluating the best model over the full dataset. Comments

The work should be submitted as a report in pdf or html format. The work there should consist of a text with graphs and tables needed to support your reasoning, the code should be in the Appendix. The report (without the Appendix) should be no longer than 10 pages.

This work implies some creative approach. In each case, you can do more than the specified minimum. Note, that you are welcome to use any techniques that are not mentioned in the course.