Pre-processing data

Handling missing data

Handling missing data: Plan

- Linear interpolation

Handling missing data: Plan

- Linear interpolation
- Modelling approach

Handling missing data: Plan

- Linear interpolation
- Modelling approach
- Using STL decomposition

Linear Interpolation

Idea

To fill in missing data we need to the restore the values so that they fit perfectly on the straight line (form an arithmetic progression),

$$\Delta y_t^{imp} = const$$

Linear Interpolation

Idea

To fill in missing data we need to the restore the values so that they fit perfectly on the straight line (form an arithmetic progression),

$$\Delta y_t^{imp} = const$$

Example:

10, NA, NA, 100

Linear Interpolation

Idea

To fill in missing data we need to the restore the values so that they fit perfectly on the straight line (form an arithmetic progression),

$$\Delta y_t^{imp} = const$$

Example:

10, NA, NA, 100

10, 40, 70, 100

Modelling approach to handle the missing data

1. Evaluate a model that allows the missing data:
   ARIMA or automatic ARIMA works fine!

Modelling approach to handle the missing data

1. Evaluate a model that allows the missing data: ARIMA or automatic ARIMA works fine!
2. Missing values of y_t can be replaced by the conditional mathematical expectation, assuming the estimated parameters of the model to be true,

$$y_t^{imp} = \mathbb{E}(y_t \mid \text{data})$$

For that Kalman filter is used

Modelling approach to handle the missing data

1. Evaluate a model that allows the missing data:
   ARIMA or automatic ARIMA works fine!
2. Missing values of y_t can be replaced by the conditional mathematical expectation, assuming the estimated parameters of the model to be true,

$$y_t^{imp} = \mathbb{E}(y_t \mid \text{data})$$

   For that Kalman filter is used

Modelling approach to handle the missing data

1. Evaluate a model that allows the missing data:
   ARIMA or automatic ARIMA works fine!
2. Missing values of y_t can be replaced by the conditional mathematical expectation, assuming the estimated parameters of the model to be true,

   $$y_t^{imp} = \mathbb{E}(y_t \mid \text{data})$$

   For that Kalman filter is used

The ability to evaluate a model on data with missing values is highly dependent on implementation

Using STL decomposition

1. We decompose the series with missing data into components:

   $y_t = \text{trend}_t + \text{seasonal}_t + \text{remainder}_t = \text{seasonal}_t + \text{deseason}_t.$

   STL restores the seasonal component without the gaps !

Using STL decomposition

1. We decompose the series with missing data into components:

   $$y_t = \text{trend}_t + \text{seasonal}_t + \text{remainder}_t = \text{seasonal}_t + \text{deseason}_t.$$

   STL restores the seasonal component without the gaps !
2. Recover the missing values of the deseasoned series by linear interpolation

Using STL decomposition

1. We decompose the series with missing data into components:

   $$y_t = \text{trend}_t + \text{seasonal}_t + \text{remainder}_t = \text{seasonal}_t + \text{deseason}_t.$$

   STL restores the seasonal component without the gaps !
2. Recover the missing values of the deseasoned series by linear interpolation
3. The missing values of y_t are replaced by the sum of the restored deseasoned values and the seasonal component,

   $$y_t^{imp} = \text{seasonal}_t + \text{deseason}_t^{imp}$$

Why we need to handle the missing data?

- Filling in the gaps is sometimes a main task

Why we need to handle the missing data?

- Filling in the gaps is sometimes a main task
- Possibility to use more algorithms of prediction for the reconstructed series

Why we need to handle the missing data?

- Filling in the gaps is sometimes a main task
- Possibility to use more algorithms of prediction for the reconstructed series
- Ability to use the restored row as a predictor

Handling missing data: Summary

- Linear interpolation: simple and fast!

Handling missing data: Summary

- Linear interpolation: simple and fast!
- Using ARIMA or more complex models

Handling missing data: Summary

- Linear interpolation: simple and fast!
- Using ARIMA or more complex models
- STL decomposition and restoring components

Handling missing data: Summary

- Linear interpolation: simple and fast!
- Using ARIMA or more complex models
- STL decomposition and restoring components
- Variations for each algorithm

Anomaly detection

Anomaly detection: Plan

- Which observation is anomalous?

Anomaly detection: Plan

- Which observation is anomalous?
- Algorithms for detection and correction of anomalies

Anomaly detection: Plan

- Which observation is anomalous?
- Algorithms for detection and correction of anomalies
- Why we look for anomalies?

Which observation is considered anomalous?

Which observation is considered anomalous?

Categorizing the observations into anomalous and ordinary one is subjective.

Which observation is considered anomalous?

Categorizing the observations into anomalous and ordinary one is subjective.

Informally, an anomalous observation doesn't fit into the normal dynamics of the series.

Which observation is considered anomalous?

Categorizing the observations into anomalous and ordinary one is subjective.

Informally, an anomalous observation doesn't fit into the normal dynamics of the series.

What is considered a "normal dynamics"? What does "not fitting in" mean?

Anomaly detection algorithm

- We take any algorithm that allows us to obtain residuals $\hat{u}_t$ from the series

Anomaly detection algorithm

- We take any algorithm that allows us to obtain residuals $\hat{u}_t$ from the series
  Residuals is the difference between the actual value and the prediction within the training set

Anomaly detection algorithm

- We take any algorithm that allows us to obtain residuals $\hat{u}_t$ from the series
  Residuals is the difference between the actual value and the prediction within the training set
  The $ARIMA$, ETS, …models, as well as the STL algorithm will do

Anomaly detection algorithm

- We take any algorithm that allows us to obtain residuals $\hat{u}_t$ from the series
  Residuals is the difference between the actual value and the prediction within the training set
  The $ARIMA$, ETS, ...models, as well as the STL algorithm will do
- Estimate standard error of the residuals

Anomaly detection algorithm

- We take any algorithm that allows us to obtain residuals $\hat{u}_t$ from the series
  Residuals is the difference between the actual value and the prediction within the training set
  The $ARIMA$, ETS, …models, as well as the STL algorithm will do
- Estimate standard error of the residuals
- If the absolute value of the remainder is greater than three standard errors, we consider the observation to be anomalous

Correction of anomalies

Correction of anomalies

We subtract the remainder from the anomalous observation:

$$y_t^{imp} = y_t - \hat{u}_t$$

Why look for anomalous sightings?

- Sometimes anomaly detection is the main goal

Why look for anomalous sightings?

- Sometimes anomaly detection is the main goal
- Possibility to get more accurate predictions for the corrected series

Why look for anomalous sightings?

- Sometimes anomaly detection is the main goal
- Possibility to get more accurate predictions for the corrected series
- Possibility to get more accurate predictions by using the corrected series as a predictor

Anomaly detection: Summary

- We take any algorithm (STL, ARIMA, ETS, …) that extracts residual from the series

Anomaly detection: Summary

- We take any algorithm (STL, ARIMA, ETS, ...) that extracts residual from the series
- There are a bunch of special algorithms

Anomaly detection: Summary

- We take any algorithm (STL, ARIMA, ETS, ...) that extracts residual from the series
- There are a bunch of special algorithms
- If the remainder is large, then we consider the observation to be anomalous

Anomaly detection: Summary

- We take any algorithm (STL, ARIMA, ETS, ...) that extracts residual from the series
- There are a bunch of special algorithms
- If the remainder is large, then we consider the observation to be anomalous
- To correct the anomalous observation, we replace the its remainder with zero

Anomaly detection: Summary

- We take any algorithm (STL, ARIMA, ETS, ...) that extracts residual from the series
- There are a bunch of special algorithms
- If the remainder is large, then we consider the observation to be anomalous
- To correct the anomalous observation, we replace the its remainder with zero
- Correction of anomalous observations before forecasting can improve forecasts!

Structural break detection

Structural break detection: Plan

- What is structural break?

Structural break detection: Plan

- What is structural break?
- Detecting single structural break

Structural break detection: Plan

- What is structural break?
- Detecting single structural break
- Detecting several structural breaks

What is considered a structural break?

What is considered a structural break?

Division of the time series into periods between structural break points is subjective.

What is considered a structural break?

Division of the time series into periods between structural break points is subjective.

Informally, at the moment of structural break the behavior (pattern) of the series changes.

What is considered a structural break?

Division of the time series into periods between structural break points is subjective.

Informally, at the moment of structural break the behavior (pattern) of the series changes.

What is considered by "changing"?

The idea of detecting a single break

- We start with a penalty function that measures heterogeneity in observations $y_a, y_{a+1}, \ldots, y_b$,

$$C(y_{a:b})$$

The idea of detecting a single break

- We start with a penalty function that measures heterogeneity in observations $y_a, y_{a+1}, \ldots, y_b$,

$$C(y_{a:b})$$

- Then we iterate over all moments $\tau \in [1; T-1]$ and find the minimum of the value

$$C(y_{1:\tau}) + C(y_{\tau+1:T})$$

The idea of detecting a single break

- We start with a penalty function that measures heterogeneity in observations $y_a, y_{a+1}, \ldots, y_b,$

$$C(y_{a:b})$$

- Then we iterate over all moments $\tau \in [1; T-1]$ and find the minimum of the value

$$C(y_{1:\tau}) + C(y_{\tau+1:T})$$

We suspect that the break could be at this point τ^*

The idea of detecting a single break

- We start with a penalty function that measures heterogeneity in observations $y_a, y_{a+1}, \ldots, y_b$,

$$C(y_{a:b})$$

- Then we iterate over all moments $\tau \in [1; T-1]$ and find the minimum of the value

$$C(y_{1:\tau}) + C(y_{\tau+1:T})$$

We suspect that the break could be at this point τ^*

- We assume that the break was in τ^* if the total heterogeneity of the the two fragments is significantly less than the heterogeneity of the entire series,

$$C(y_{1:\tau^*}) + C(y_{\tau^*+1:T}) < C(y_{1:T}) - \beta$$

Choice of penalty function C

- There are many options

Choice of penalty function C

- There are many options
- Often we can take the log-likelihood function of some model, multiplied by minus two:

$$C(y_{a:b}) = -2 \max_{\theta} \ln L(y_a, \ldots, y_b \mid \theta)$$

Choice of penalty function C

- There are many options
- Often we can take the log-likelihood function of some model, multiplied by minus two:

$$C(y_{a:b}) = -2 \max_{\theta} \ln L(y_a, \ldots, y_b \mid \theta)$$

The simplest model: $y_t \sim \mathcal{N}(\mu, \sigma^2)$ and are independent

Choice of penalty function C

- There are many options
- Often we can take the log-likelihood function of some model, multiplied by minus two:

$$C(y_{a:b}) = -2 \max_{\theta} \ln L(y_a, \ldots, y_b \mid \theta)$$

The simplest model: $y_t \sim \mathcal{N}(\mu, \sigma^2)$ and are independent

- The choice of the C function is related to the choice of β when checking for a break at the assumed break point τ^*,

$$C(y_{1:\tau^*}) + C(y_{\tau^*+1:T}) < C(y_{1:T}) - \beta$$

Choice of penalty function C

- There are many options
- Often we can take the log-likelihood function of some model, multiplied by minus two:

$$C(y_{a:b}) = -2 \max_{\theta} \ln L(y_a, \ldots, y_b \mid \theta)$$

  The simplest model: $y_t \sim \mathcal{N}(\mu, \sigma^2)$ and are independent
- The choice of the C function is related to the choice of β when checking for a break at the assumed break point τ^*,

$$C(y_{1:\tau^*}) + C(y_{\tau^*+1:T}) < C(y_{1:T}) - \beta$$

  The more parameters in θ, the larger β should be

How to detect many structural breaks?

- Run algorithm to detect single structural break

How to detect many structural breaks?

- Run algorithm to detect single structural break
  If the algorithm did not detect a break, then we consider that there are no structural break in the series

How to detect many structural breaks?

- Run algorithm to detect single structural break
  If the algorithm did not detect a break, then we consider that there are no structural break in the series
- Else divide the original series into two sections according to the detected structural break

How to detect many structural breaks?

- Run algorithm to detect single structural break
  If the algorithm did not detect a break, then we consider that there are no structural break in the series
- Else divide the original series into two sections according to the detected structural break
- Then recursively run the detection algorithm for one structural break to each of the detected subsections

Transformations before the search

The structural break can be easier to detect on the transformed series

- Simple transformations of the initial series: logarithm, Box-Cox transformation, transition to differences

Transformations before the search

The structural break can be easier to detect on the transformed series

- Simple transformations of the initial series: logarithm, Box-Cox transformation, transition to differences
- Decomposition of the series: STL, ETS, $\ldots$

Why look for structural breaks?

- Sometimes structural break detection is the main goal

Why look for structural breaks?

- Sometimes structural break detection is the main goal
- Ability to get more accurate forecasts if a dummy variable (equal to one after the break) is added to the set of predictors

Why look for structural breaks?

- Sometimes structural break detection is the main goal
- Ability to get more accurate forecasts if a dummy variable (equal to one after the break) is added to the set of predictors
- Possibility to get more accurate forecasts of other series if corrected for the structural break series is used as the predictor

Structural break detection: Summary

- There are many specialized algorithms

Structural break detection: Summary

- There are many specialized algorithms
- Does the sum of inhomogeneities on the left and right sections to the possible break strongly differ from the heterogeneity of the entire series?

Structural break detection: Summary

- There are many specialized algorithms
- Does the sum of inhomogeneities on the left and right sections to the possible break strongly differ from the heterogeneity of the entire series?
- To find several breaks, it is enough to search for the next break in the already identified subsections of the series

Structural break detection: Summary

- There are many specialized algorithms
- Does the sum of inhomogeneities on the left and right sections to the possible break strongly differ from the heterogeneity of the entire series?
- To find several breaks, it is enough to search for the next break in the already identified subsections of the series
- STL decomposition allows you to search for breaks in the components of a series