Faculty of Computer Science

Master of Data Science

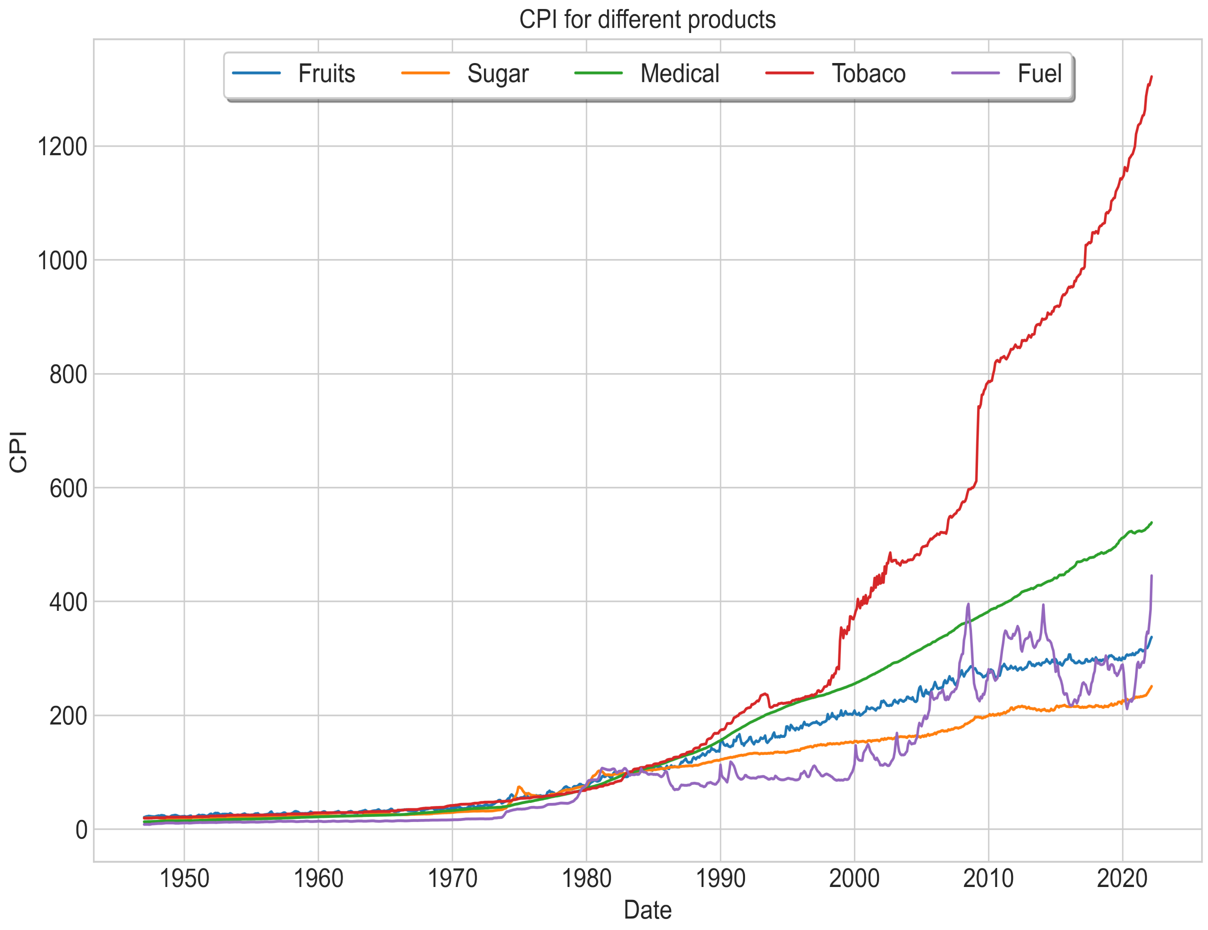
Moscow
2022

Structural break detection

Student: Rudenko Aleksey Victorovich
Supervisor: MOEX, HSE Lecturer Lukyanchenko Petr Pavlovich

Structural break

Key assumption in forecasting of time series are model stability of parameters – assumption that the predictive ability of our models over a sample period and forecast horizon does not decreasing. But in real world forecasting tasks it is not true, parametric models often change in different ways. Economy, political situation, world crisis, epidemics can lead to changing data structural, also the relationship between implied variables sometimes change.



The Chow test

$$F = \frac{(RSS-(RSS_1+RSS_2))/k}{(RSS_1+RSS_2)/(T-2k)}$$

K – number of observations in each model

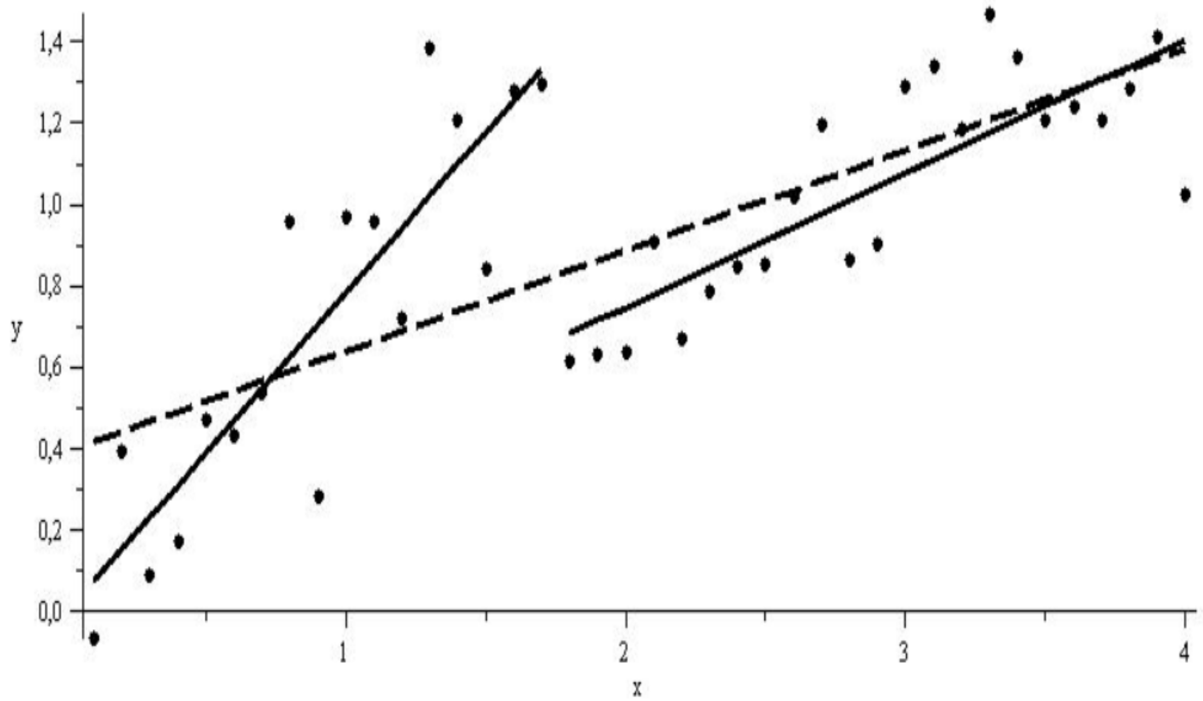
RSS – sum of squared residuals of unrestricted model

$RSS_{1,2}$ – sum of squared residuals of restricted models

T – number of all observations

H_0 : *Coefficients constant across sample (no break)*

H_1 : *Coefficients differ across sample (there is break)*

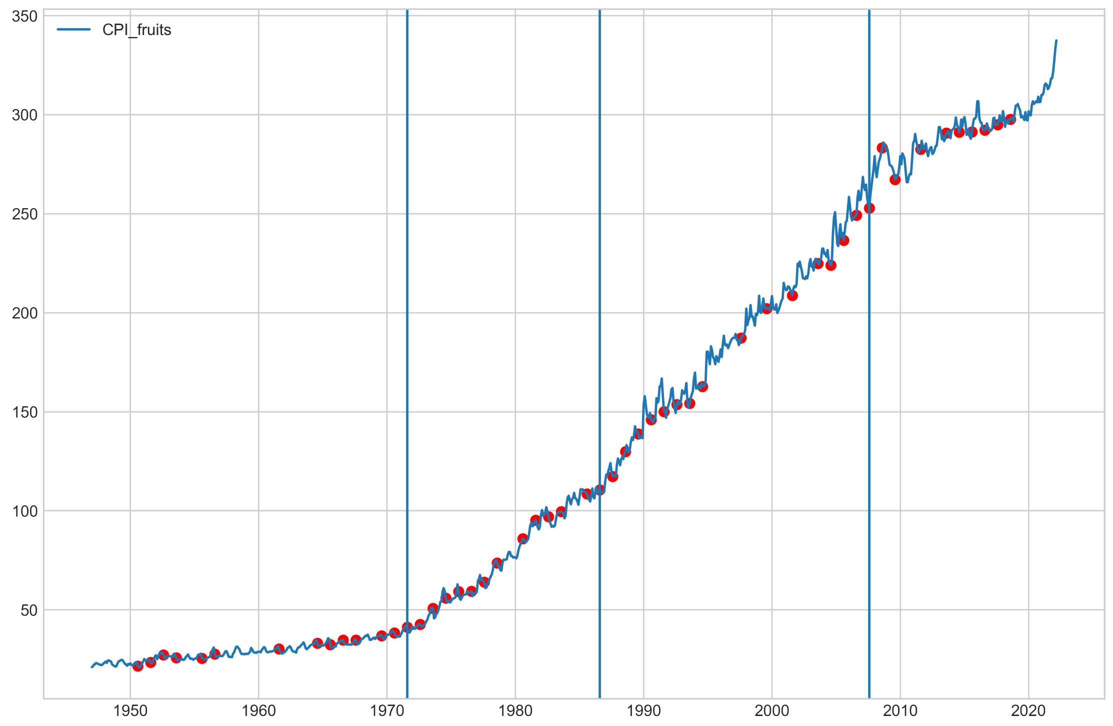


The Chow test(the break date is unknow)
The Quandt Likelihood Ratio Test

The QLR test statistic is the maximum of all the Chow F-statistics, over a range of τ, τ0 ≤ τ ≤ τ1 :

QLR = max[F(τ0), F(τ0+1) ,…, F(τ1–1), F(τ1)]

Size of sliding window: 7 – 11 years, step = 1 year

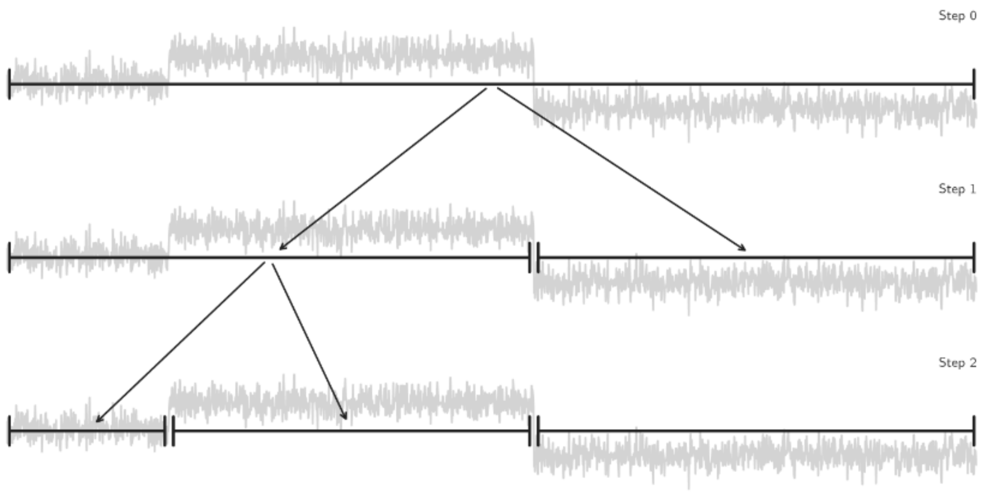


Single change point detection

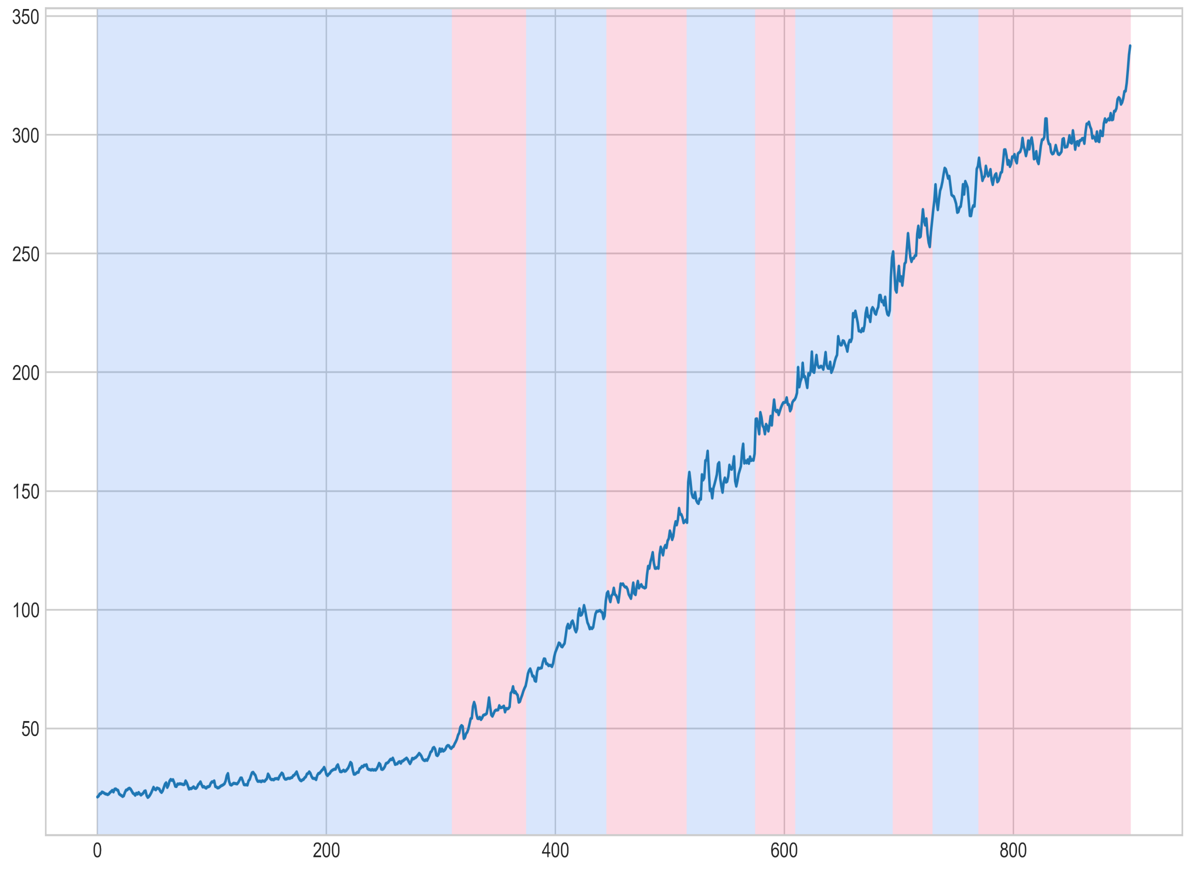
Change point is then said to arise when there exists a time, τ∈{1,…,T−1}τ∈{1,…,T−1}, such that the statistical properties of {y1,…,yτ}{y1,…,yτ} and {yτ+1,…,yT}{yτ+1,…,yT} are different

H0:no change point

H1:there is a single point



Schematic view of the binary segmentation algorithm



Forecasting time series: Exp.smoothing

Exp.smoothing: $l_t = a * y_t + (1- a) * l_{t-1}, l_0 = y_o$

Forecast: $y_{t+1} = l_t$

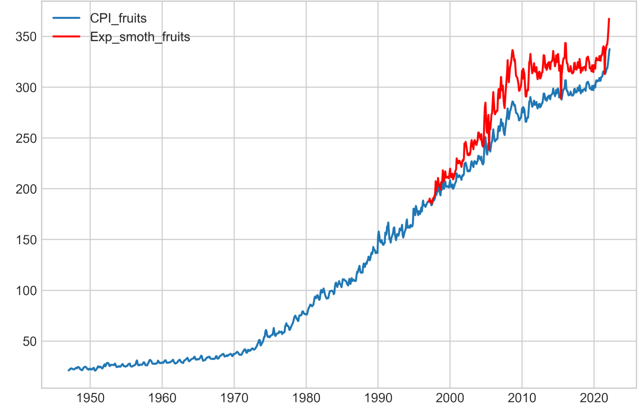
Holt-Winters(triple exp.smoothing):

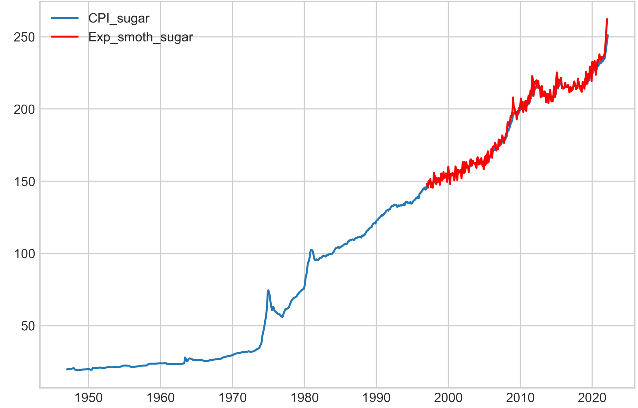
$l_t = a * (y_t - s_{t-L}) + (1- a) * (l_{t-1} + b_{t-1})$

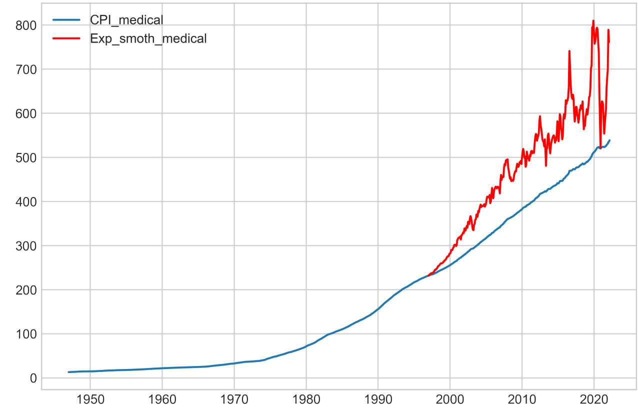
$b_t = \beta * (l_t - l_{t-1}) + (1-\beta) * (b_{t-1})$

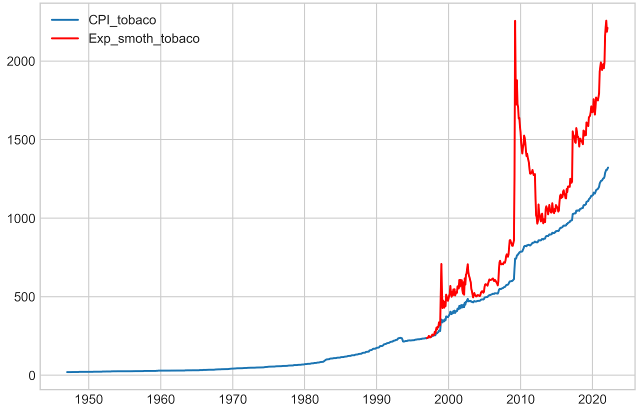
$s_t = \gamma * (y_t - l_t) + (1-\gamma) * (s_{t-L})$

$y_{t+1} = l_t + m * b_t + (s_{t-L})$





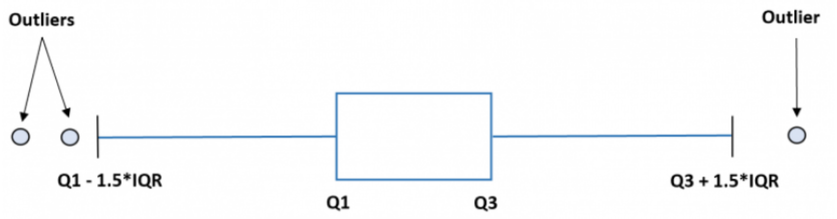


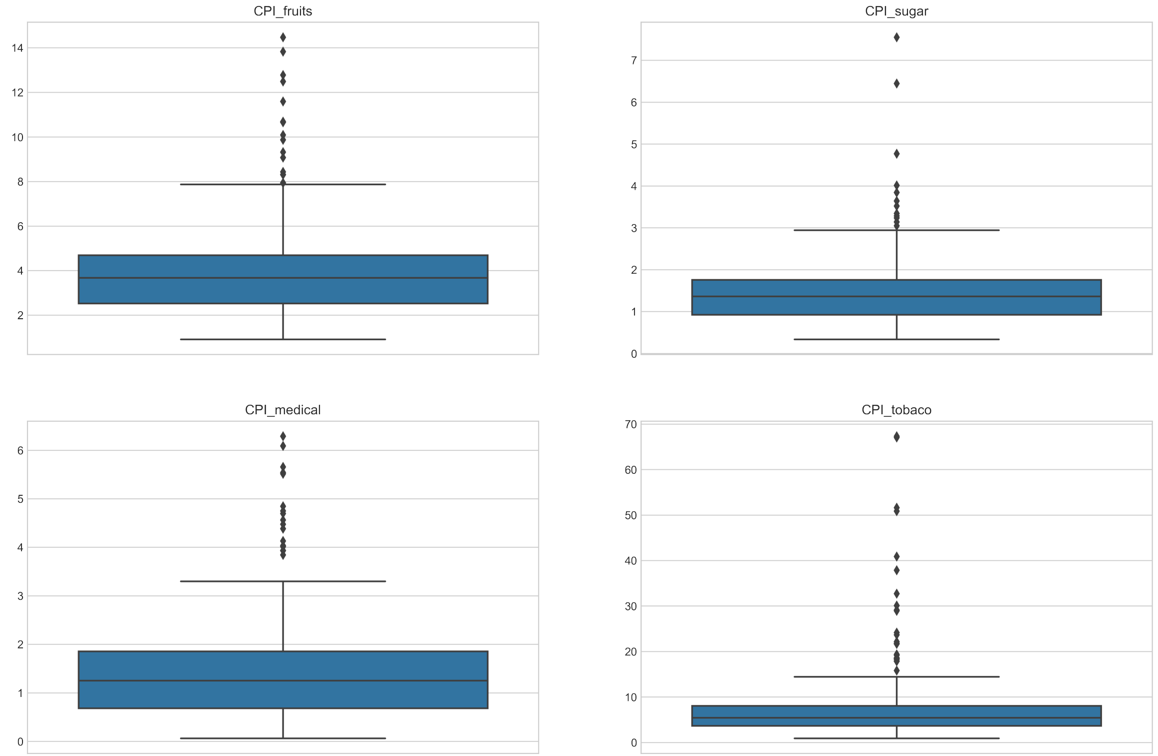


Choosing outliers

Lower limit = Q1 – 1.5*IQR

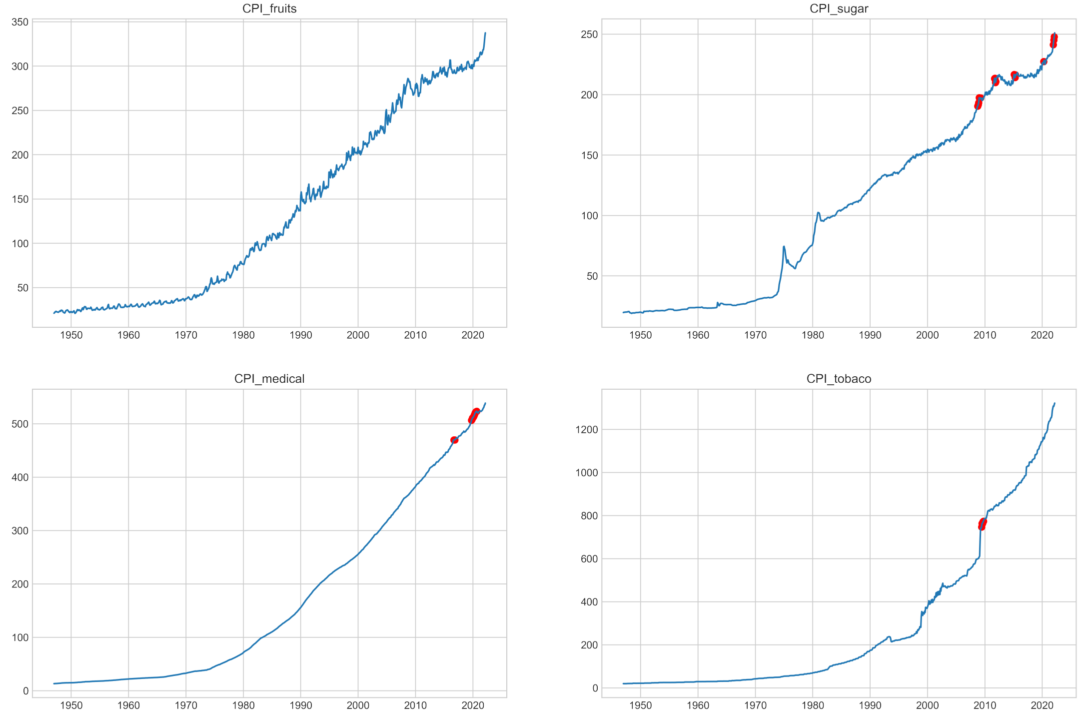
Upper limit = Q3 + 1.5*IQR

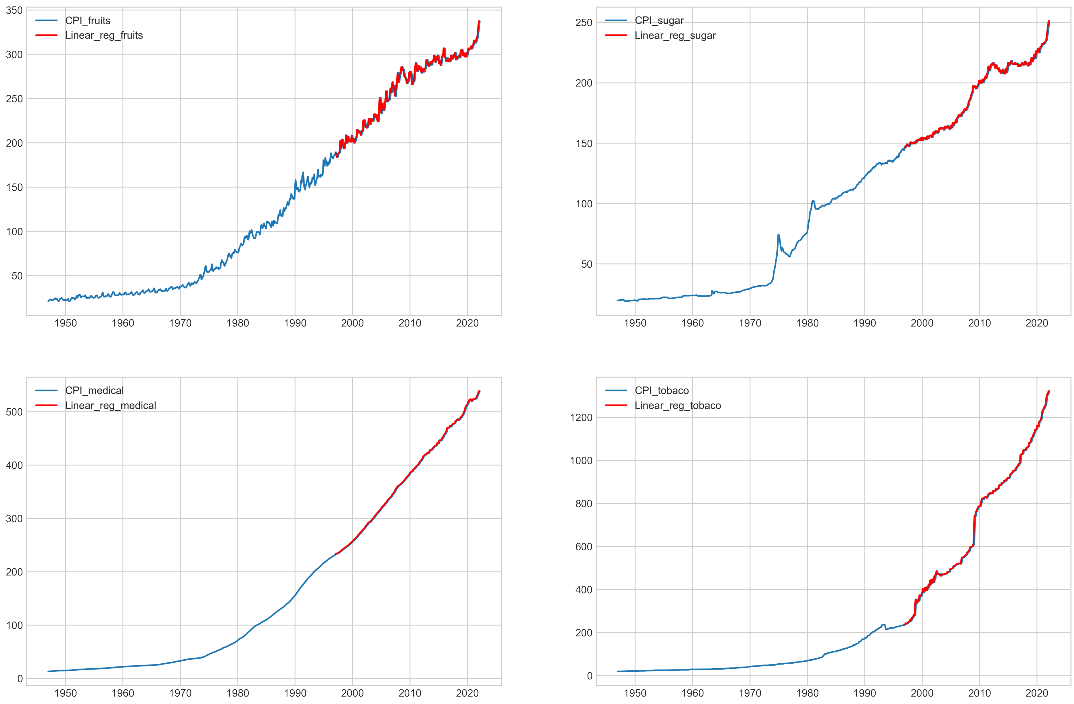




Forecasting time series: Linear regression

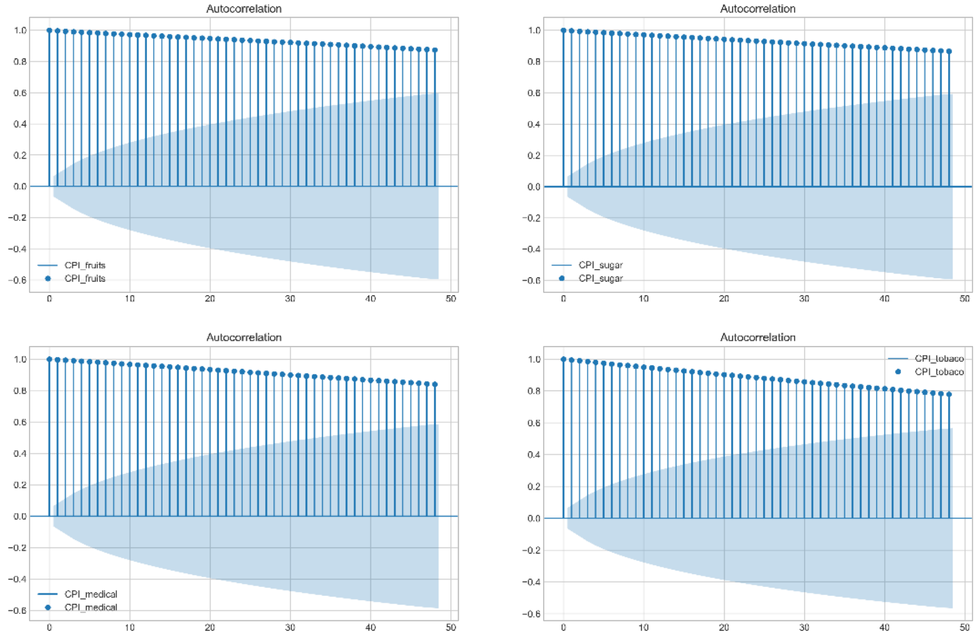
Features: Lag feature(from 1 – 6), deterministic processes variable

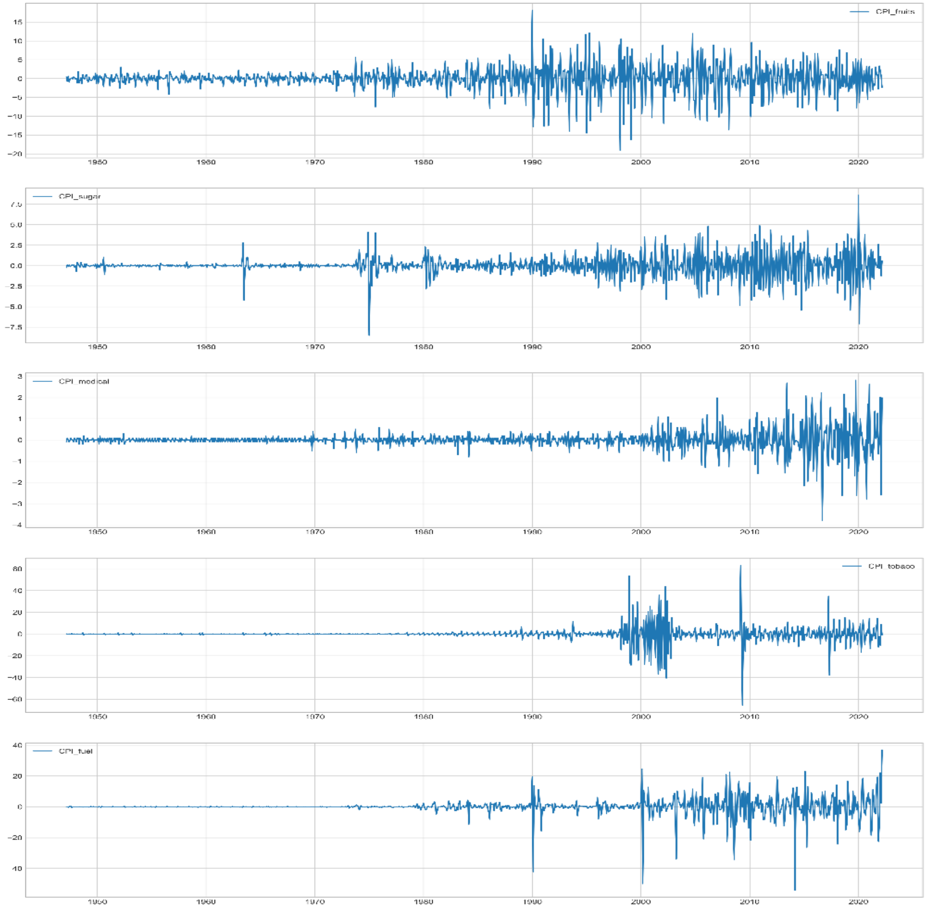




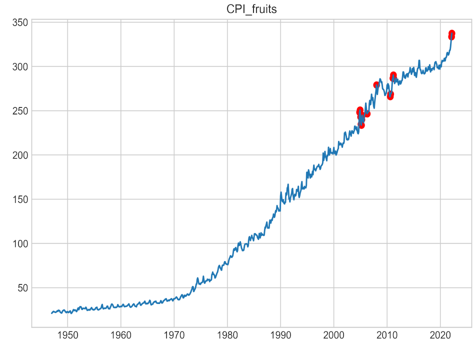
Forecasting time series: ARIMA

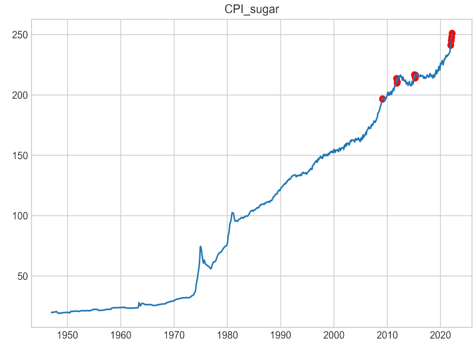
$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

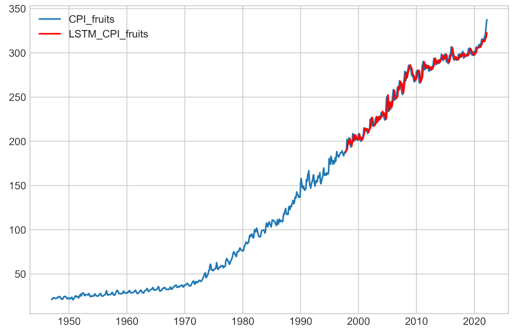


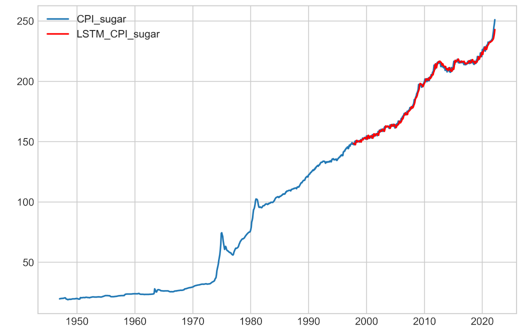


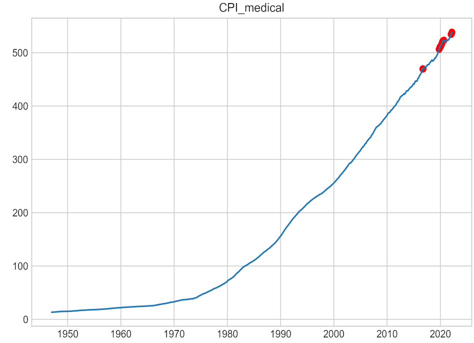
Forecasting time series: LSTM

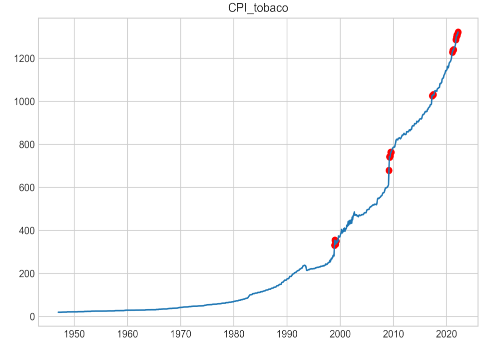


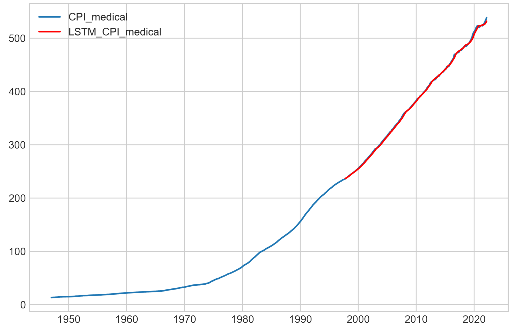


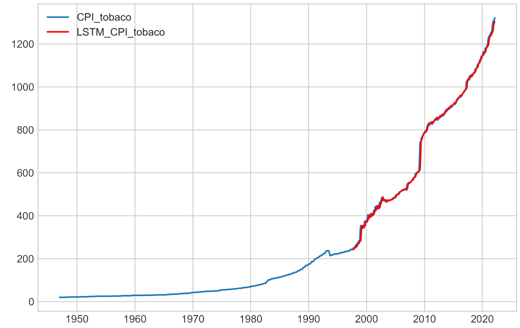




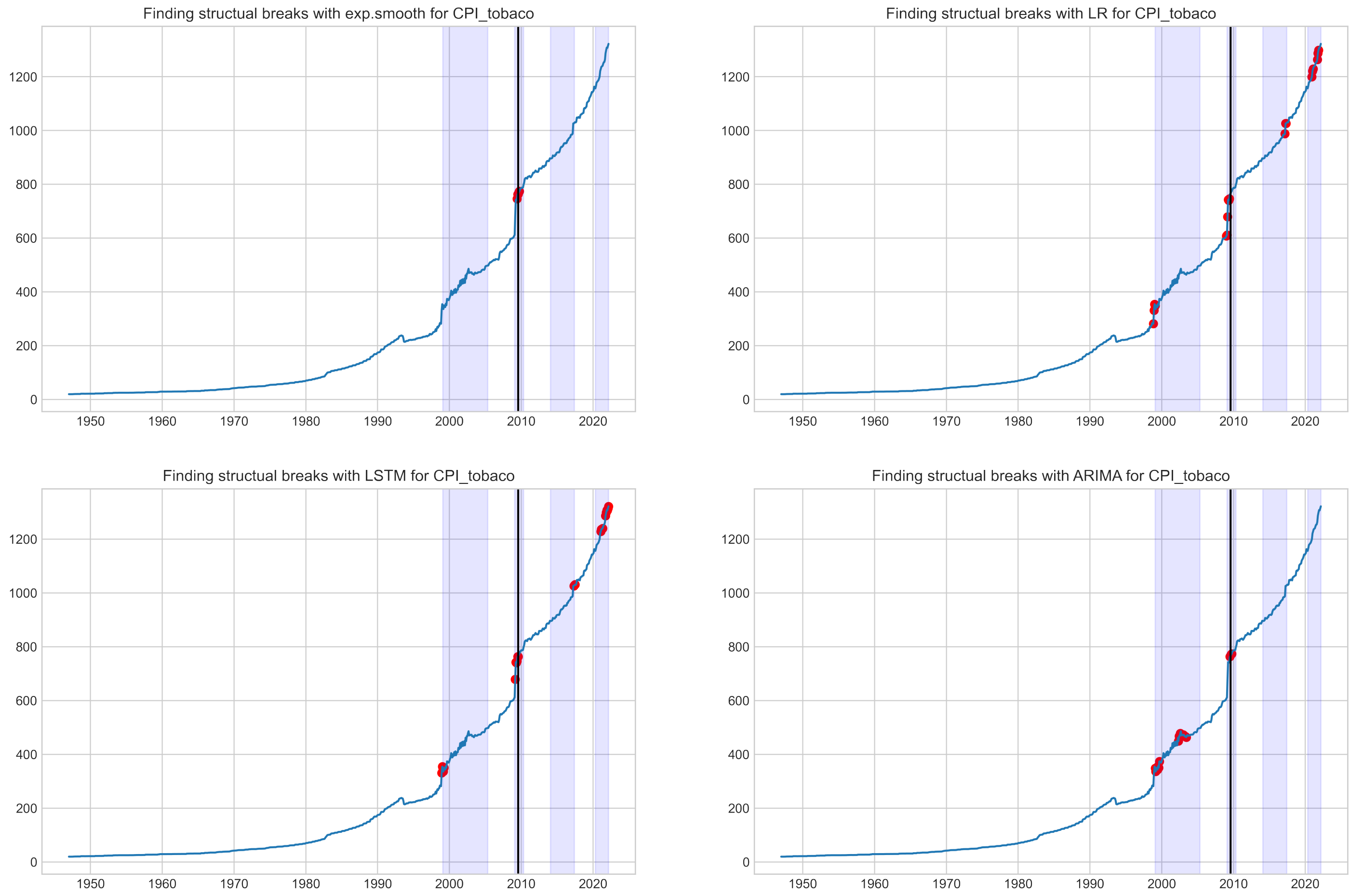








Comparing results



Conclusions

- Classical methods for detecting points of structural breaks have been considered
- Different(exp.smoothing, LR, ARIMA, LSTM) time series forecasting models were trained
- Considered a method for finding structural break points based on model forecast errors

THANK YOU
FOR YOUR
ATTENTION