TIME SERIES AND FORECASTING: FINAL PROJECT

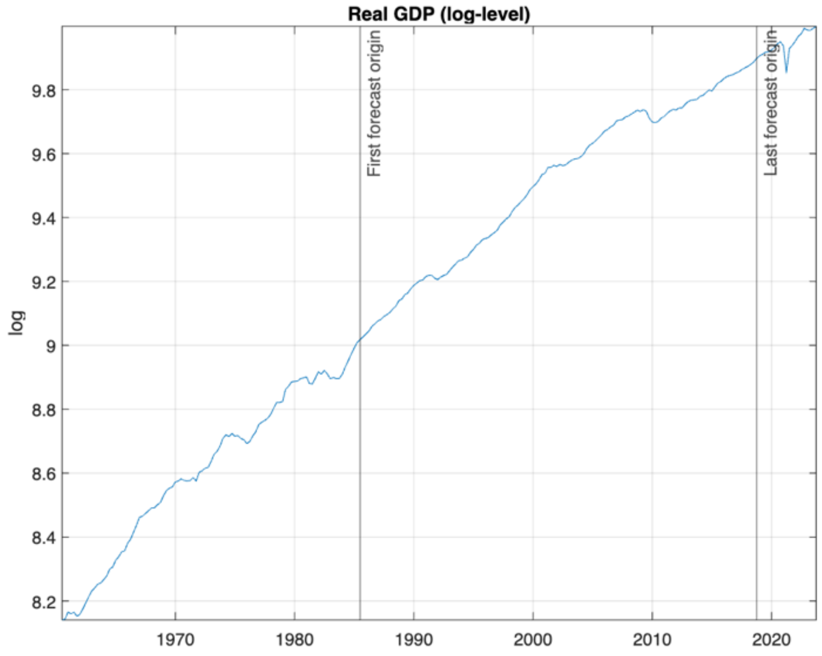
Beatrice CAGNIN, matricola 13626A
Jing WANG, matricola 13347A

DATA DESCRIPTION

- **GDPC1** = Real Gross Domestic Product
- **PCECTPI** = Personal Consumption Expenditure
- **TB3MS** = 3-Month Treasury Bill Secondary Market Rate (Percent)
- **GS10** = 10-Year Treasury Constant Maturity Rate (Percent)

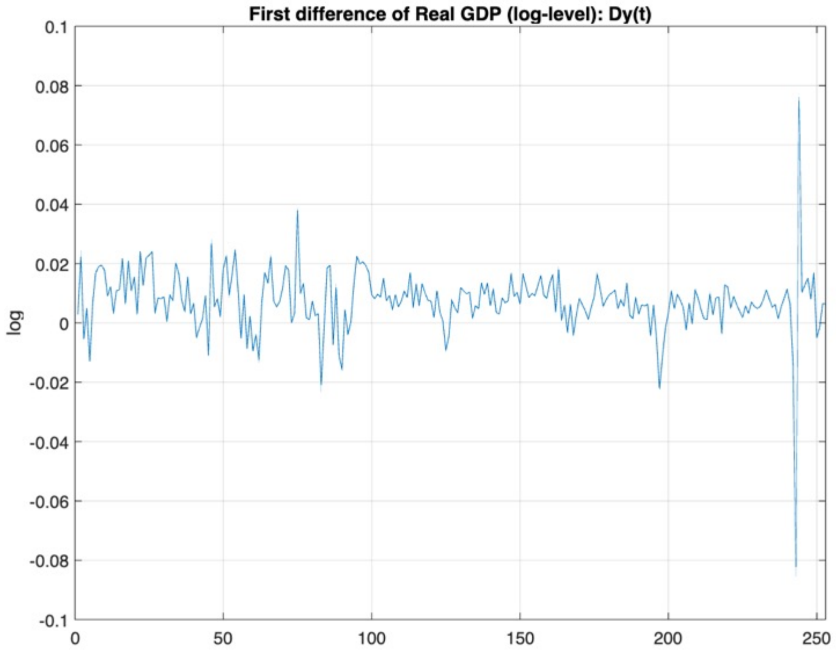
1) PLOTS OF SERIES USED IN THE ANALYSIS

Y_t is the log of GDPC1



It displays an upward trend.

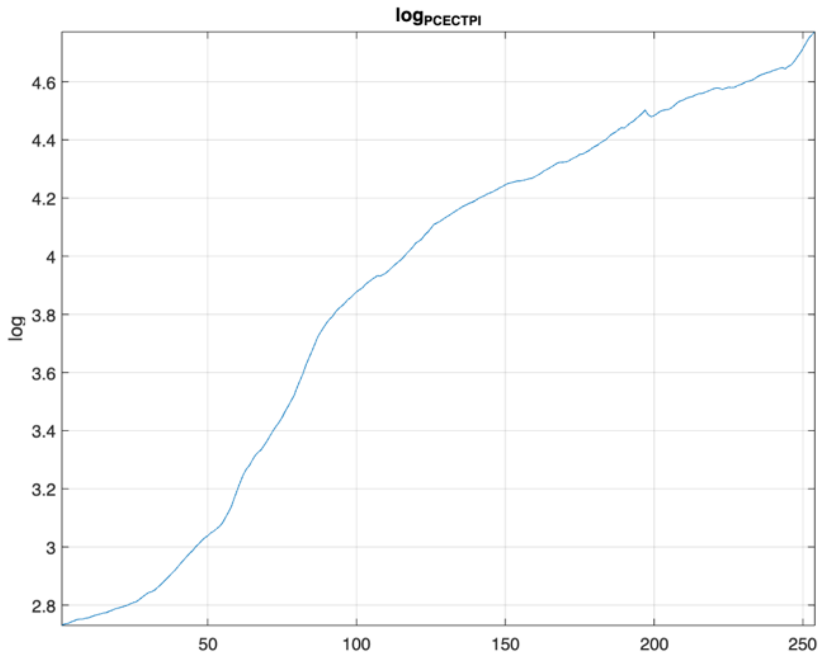
ΔY_t is the first difference of Y_t



The trend disappears and there's mean reversion around 0.01.

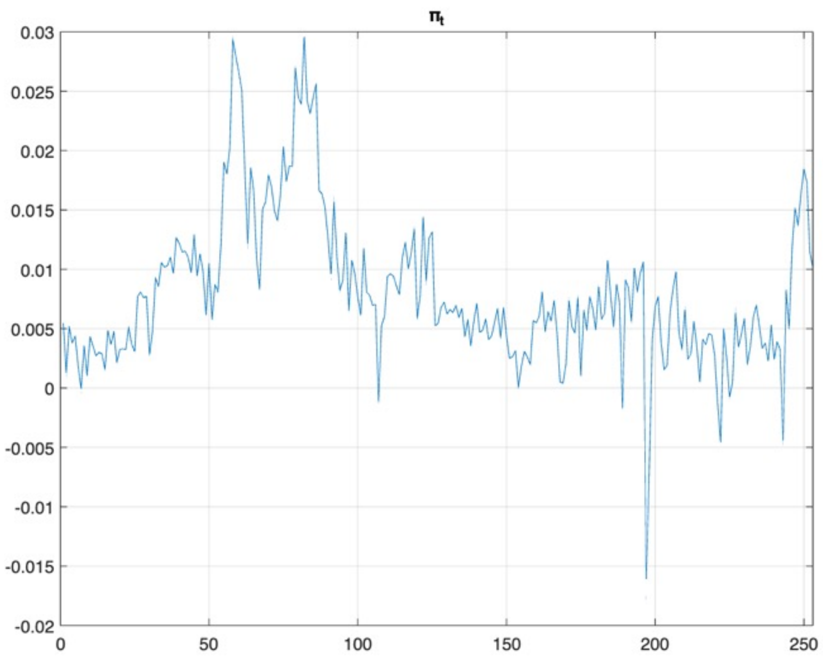
1) PLOTS OF SERIES USED IN THE ANALYSIS

logPCECTPI_t is the log of PCECTPI



It displays an upward trend.

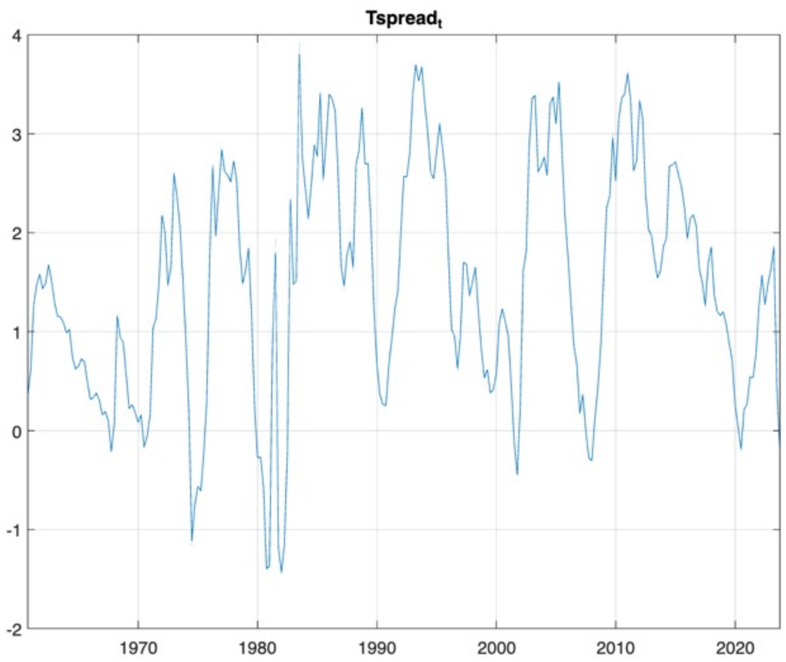
π_t is the first difference of logPCECTPI_t



The trend disappears and there can be a cycle in the long run.

1) PLOTS OF SERIES USED IN THE ANALYSIS

Tspread_t is the difference between GS10 and TB3MS

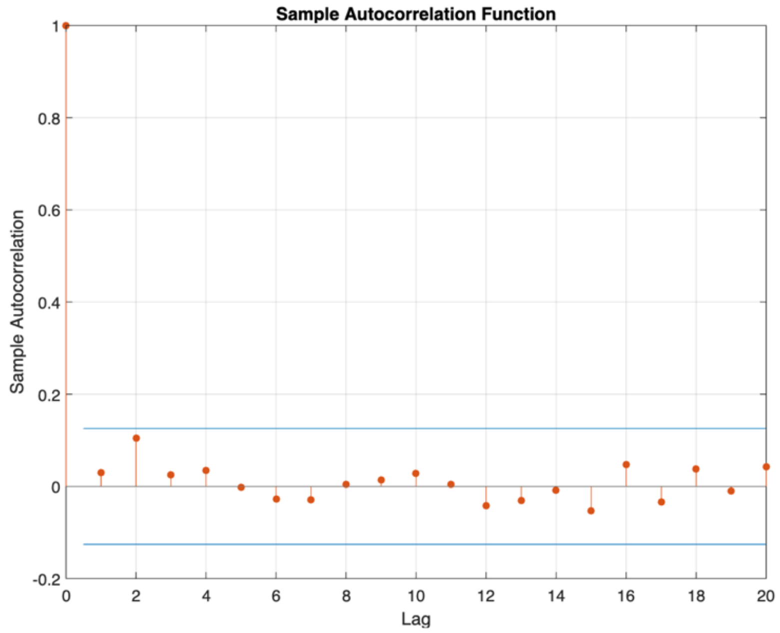


It displays no trend and some mean reversion.

2) VAR(4) MODEL

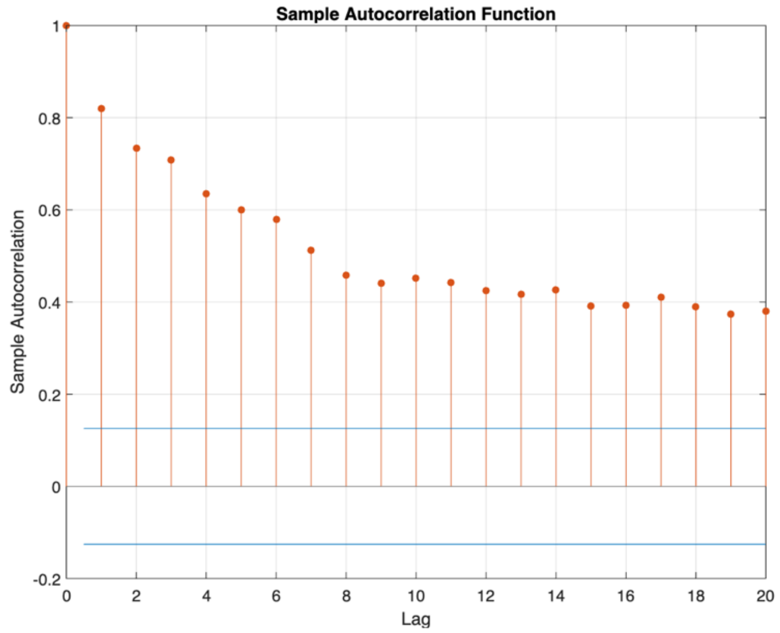
Sample Autocorrelation functions of the series entering the VAR(4) model:

Sample ACF of Δy_t



The trend has been removed: all the spikes are within the confidence bands at 95%.

Sample ACF of π_t

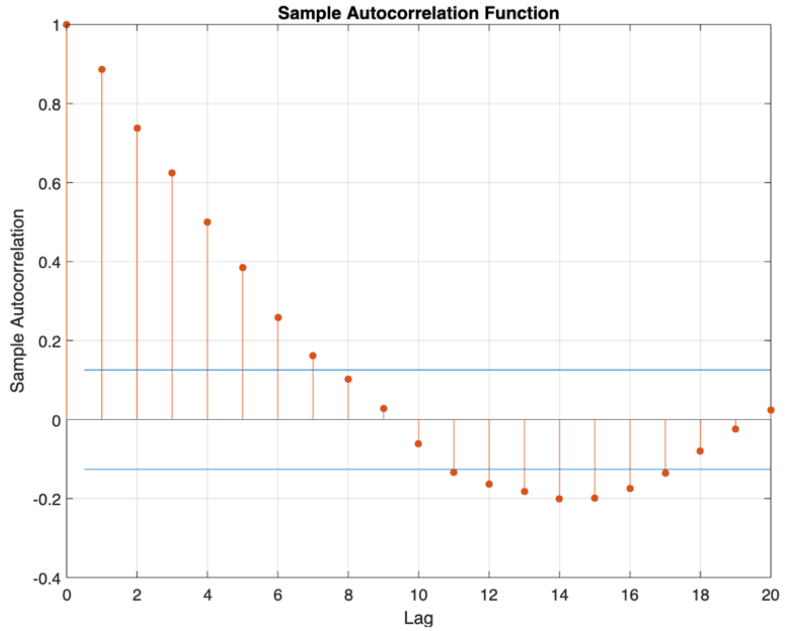


All the spikes are significant until lag=20, so we reject H_0 of no serial correlation.
Being a 1st diff, the trend remains, so there's no stationarity.

2) VAR(4) MODEL

Sample Autocorrelation functions of the series entering the VAR(4) model:

Sample ACF of $Tspread_t$



There are significant spikes until lag=7, and also from lag=11 to lag=17 where the autocorrelation is negative.

2) VAR(P) MODEL

VAR(p) model using the AIC to find the optimal n°lags

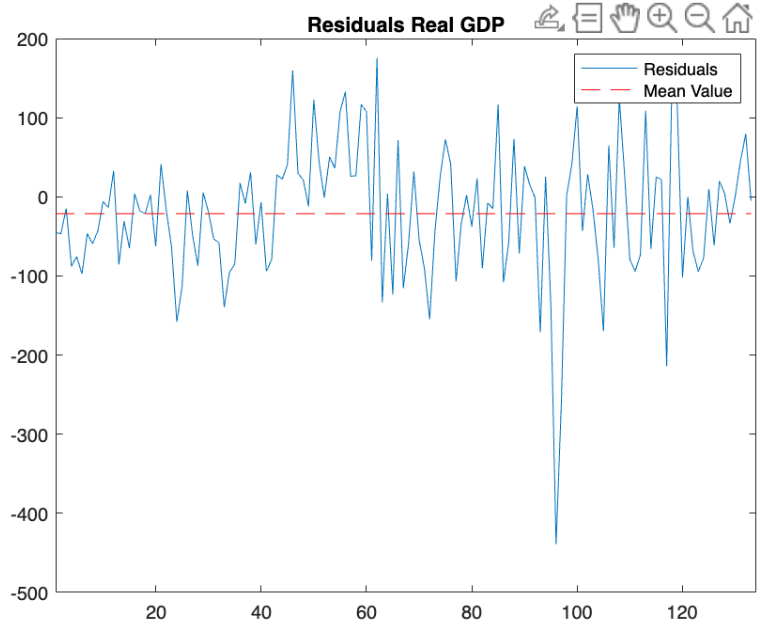
```
IC = NaN (12, 1);
for ii = 1:12
    mhat = estimate(varm(3, ii), y_var);
    tmp = summarize(mhat);
    IC(ii) = tmp.AIC;
end;

paic = find(min(IC) == IC) % it gives the minimum of AIC
```

```
paic = 2
```

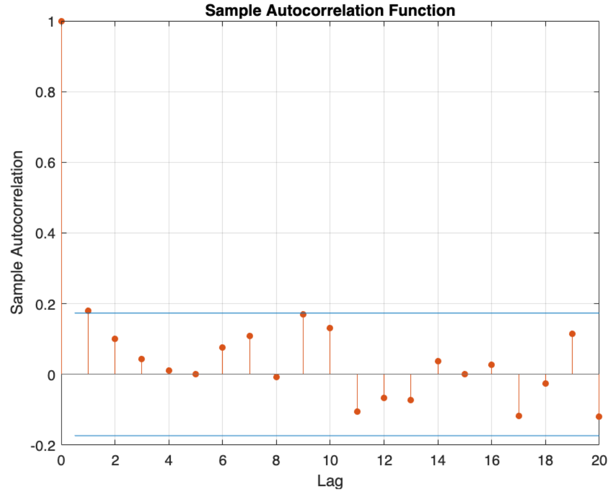
So the optimal number of lags for this model is 2.

2) RESIDUALS OF THE SERIES WITH ACF, PACF

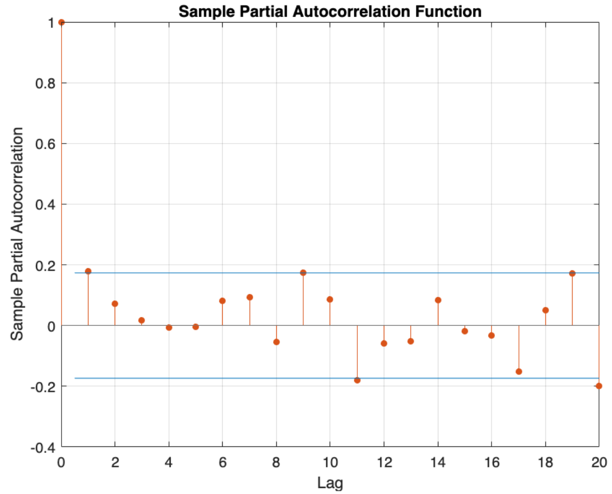
Residuals: Real GDP



Sample ACF



Sample PACF

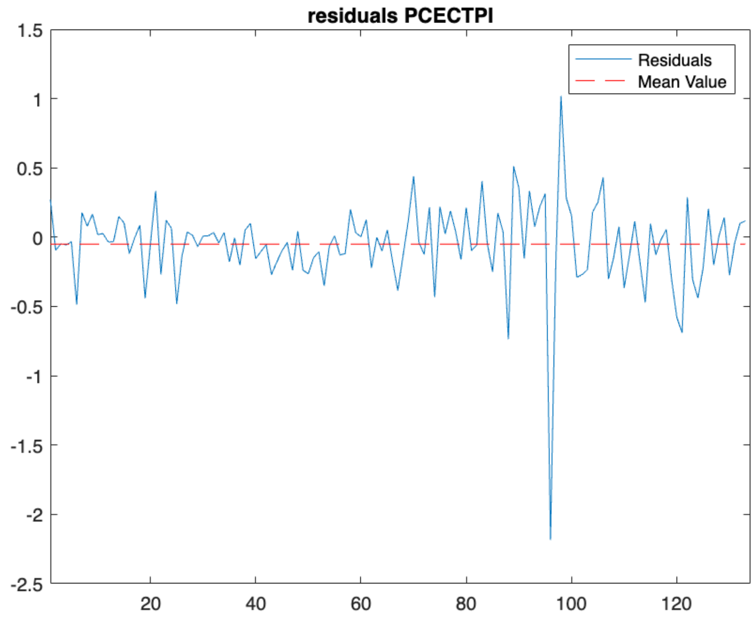


Apart from a huge spike, the residuals are between -200 and +200.

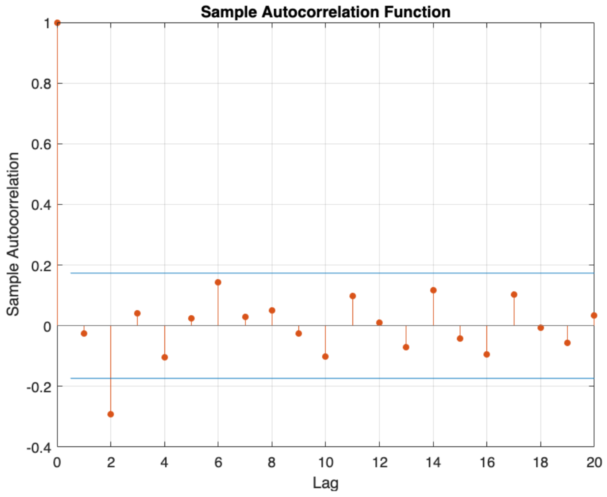
ACF and PACF display no significant lags.

2) RESIDUALS OF THE SERIES WITH ACF, PACF

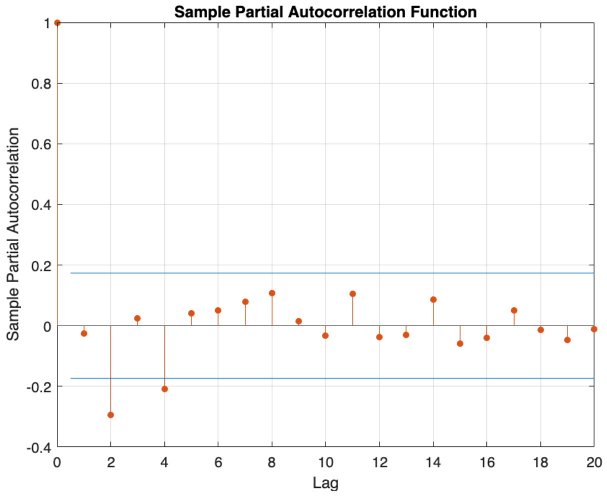
Residuals: PCECTPI



Sample ACF



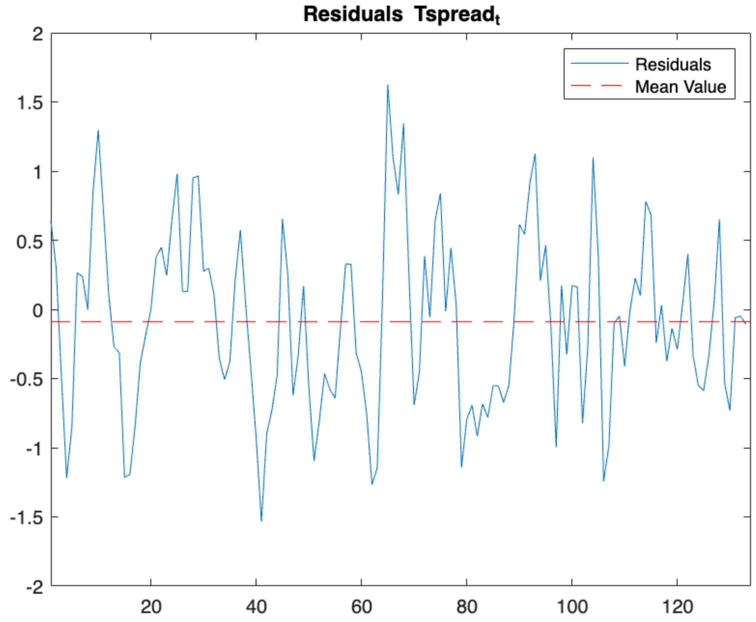
Sample PACF



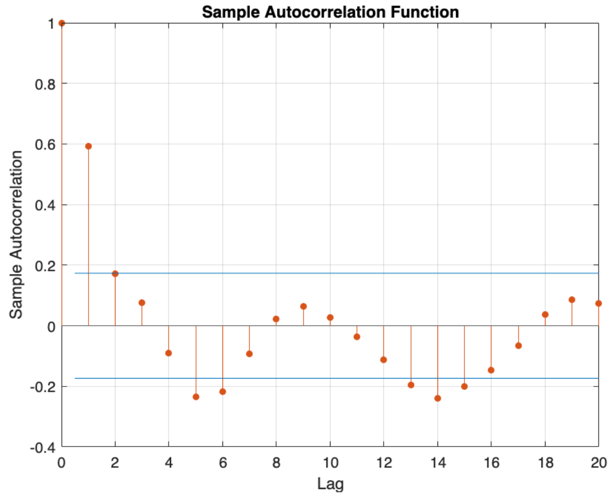
The residuals are between -2 and +1.
ACF displays a negative significant spike at lag=2.
PACF displays two negative significant spikes at lags=2,4.
Negative spikes explain overfitting in the estimation.

2) RESIDUALS OF THE SERIES WITH ACF, PACF

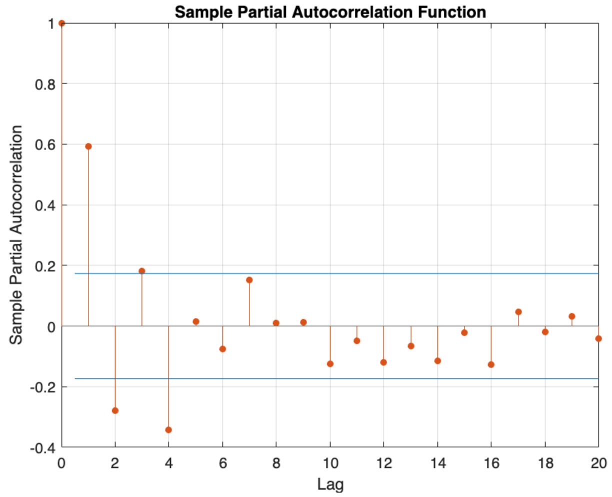
Residuals: T-spread



Sample ACF



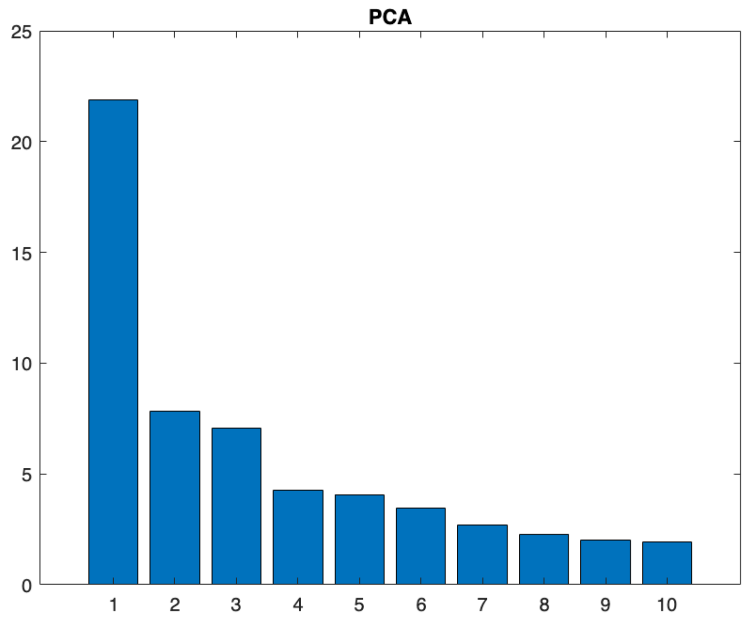
Sample PACF



The residuals are between -1.5 and +1.5.
ACF displays a positive spike at lag=1 and negative significant spikes at lags=5,6,13,14,15.
PACF displays significant spikes until lag=4.

3) PC FACTORS

PC factors obtained with PCA:



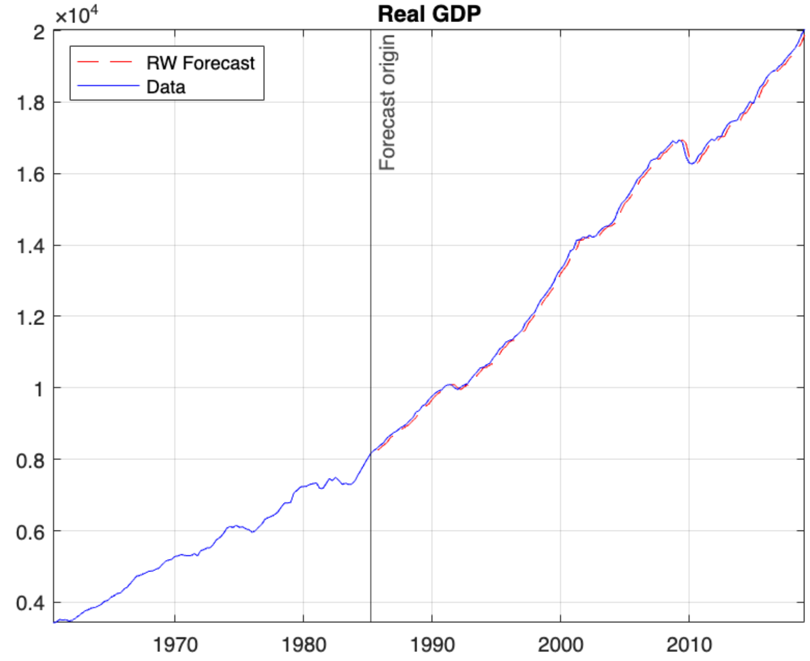
Interpretation:

The 1st factor explains the 22% of the variability of the X exogenous variables that we used to forecast.

4) ESTIMATED MODELS AND FORECASTS:

1. Random Walk

Real GDP Vs RW forecast



<u>Model</u>

$y_t = y_{t-1} + \epsilon_t$ with $\epsilon_t \sim N(0, \sigma^2)$

(is equivalent to specify a zero-mean arima model for returns: ARIMA(0,1,0))

<u>Forecast</u>

$\hat{y}_{T+h|T} = E(y_{T+h}|T) = y_T$ with $h = 1$

$RMSE_{RW} = 120.0598$

4) ESTIMATED MODELS AND FORECASTS:

2. AR(4)

Model

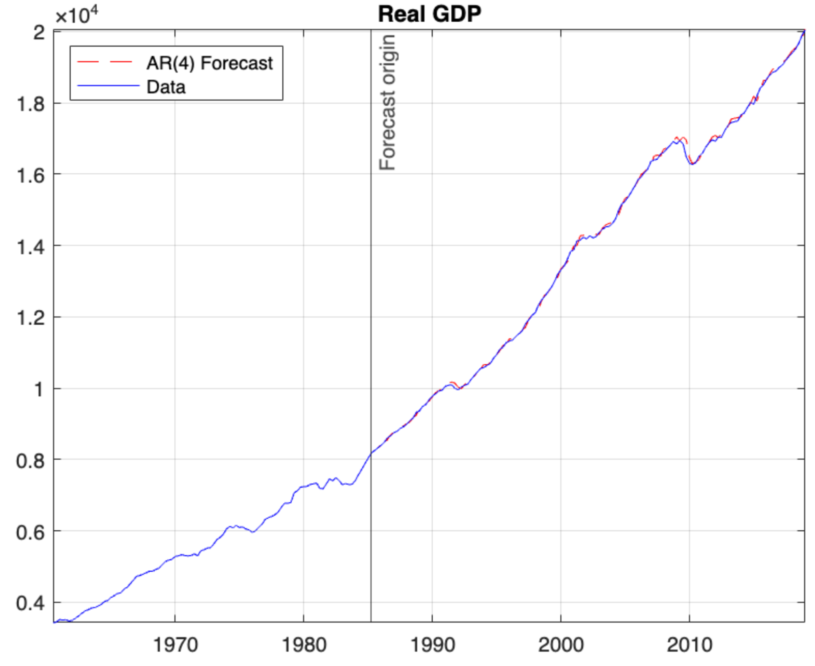
$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \rho_4 y_{t-4} + \epsilon_t$$

with $\epsilon_t \sim WN(0, \sigma^2)$

Forecast

using ARIMA(4, 1, 0) because we are considering the first difference

Real GDP Vs AR(4) forecast



$$RMSE_{AR(4)} = 80.3816$$

4) ESTIMATED MODELS AND FORECASTS:

3. VAR(4)

Model (with companion form)

$$y_t = A_c y_{t-1} + \epsilon_t \quad \text{with } \epsilon_t \sim WN(0, \Sigma_\epsilon)$$

Estimation using varm(3, 4) model

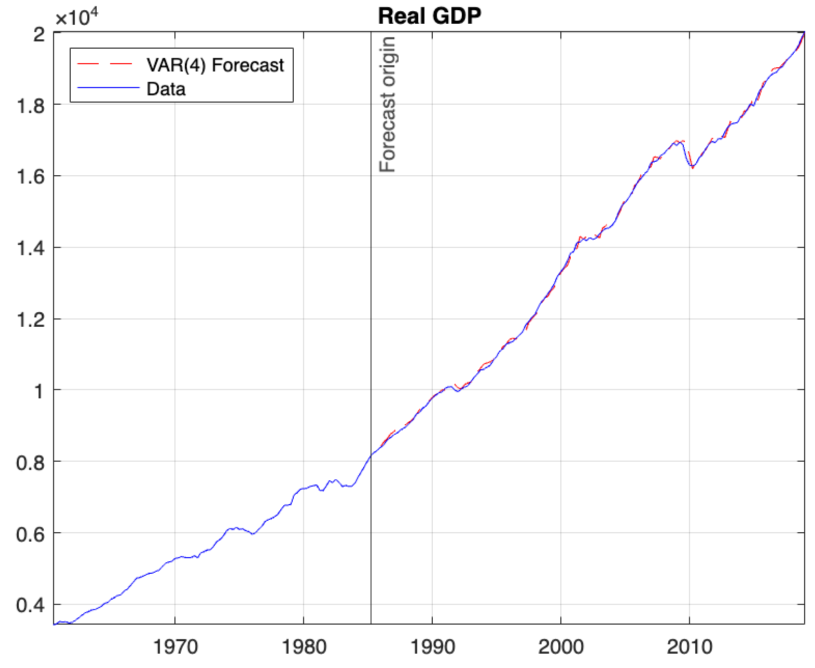
Forecast

$$\hat{y}_{T+h|T} = \hat{A}_c^h \hat{y}_{T+h-1|T} = \hat{A}_c^h \hat{y}_T$$

Forecast using

$$\Delta\hat{y}_{T+1|T} = \hat{y}_{T+1|T} - y_T \leftrightarrow \hat{y}_{T+1|T} = y_T + \Delta\hat{y}_{T+1|T}$$

Real GDP Vs VAR(4) forecast



$$RMSE_{VAR(4)} = 92.9077$$

4) ESTIMATED MODELS AND FORECASTS:

4. VAR(2)

Model (with companion form)

$$y_t = A_c y_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \sim WN(0, \Sigma_\epsilon)$$

Estimation using varm(3, 2) model

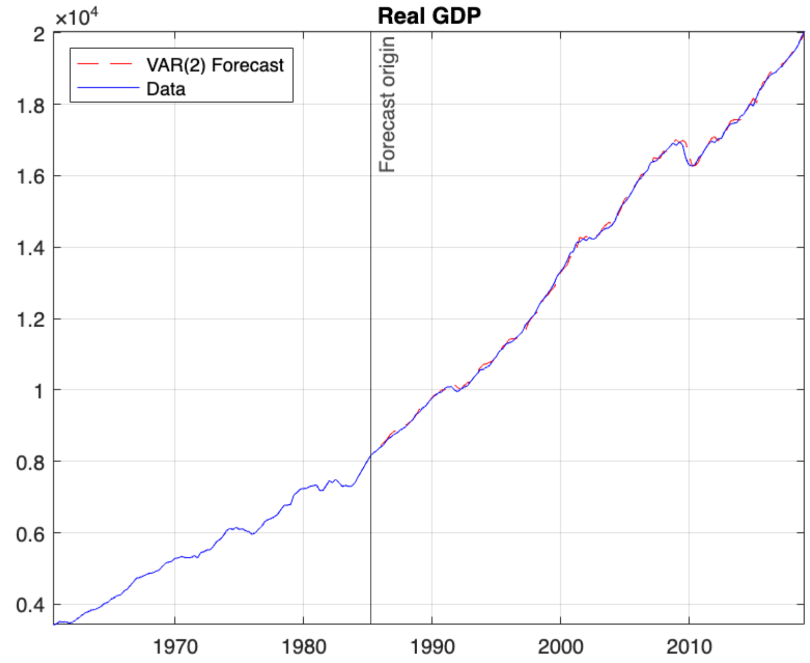
Forecast

$$\hat{y}_{T+h|T} = \hat{A}_c^h \hat{y}_{T+h-1|T} = \hat{A}_c^h \hat{y}_T$$

Forecast using

$$\Delta\hat{y}_{T+1|T} = \hat{y}_{T+1|T} - y_T \leftrightarrow \hat{y}_{T+1|T} = y_T + \Delta\hat{y}_{T+1|T}$$

Real GDP Vs VAR(2) forecast

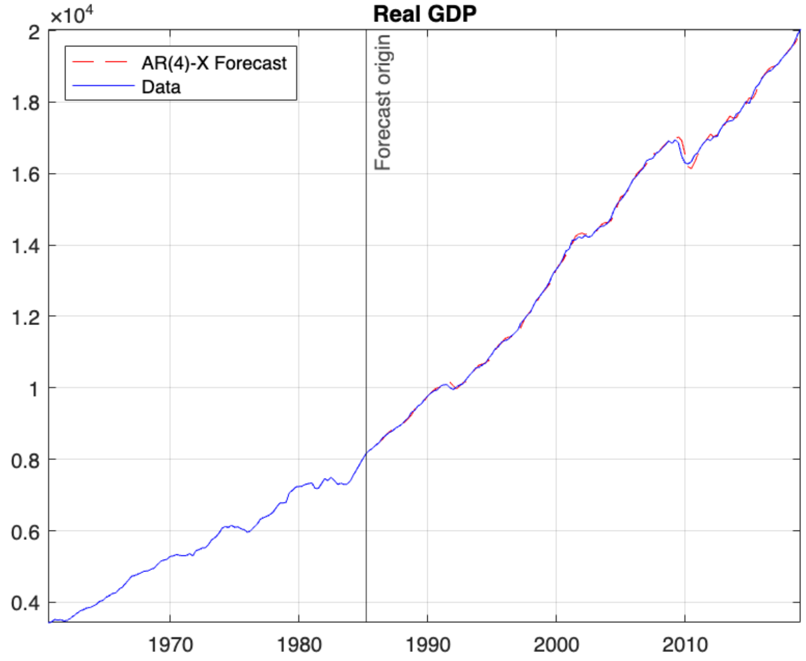


$$RMSE_{VAR(2)} = 87.9803$$

4) ESTIMATED MODELS AND FORECASTS:

4. AR(4)-X

Real GDP Vs AR(4)-X forecast



<u>Model</u>

$y_t = \sum_{j=1}^{4} \rho_j y_{t-j} + \beta F_{1,t} + \epsilon_t$ with $h = 1$

Estimation using $\Delta y_t = \sum_{j=1}^{4} \rho_j \Delta y_{t-j} + \beta F_{1,t-1}$

(we used only the first factor for the estimation as required).

<u>Forecast</u>

$$\Delta\hat{y}_{T+1|T} = \hat{y}_{T+1|T} - y_T \leftrightarrow \hat{y}_{T+1|T} = y_T + \Delta\hat{y}_{T+1|T}$$

$$RMSE_{AR(4)-X} = 83.9546$$

4) ESTIMATED MODELS AND FORECASTS:

5. BONUS: Random Walk with Drift

Model

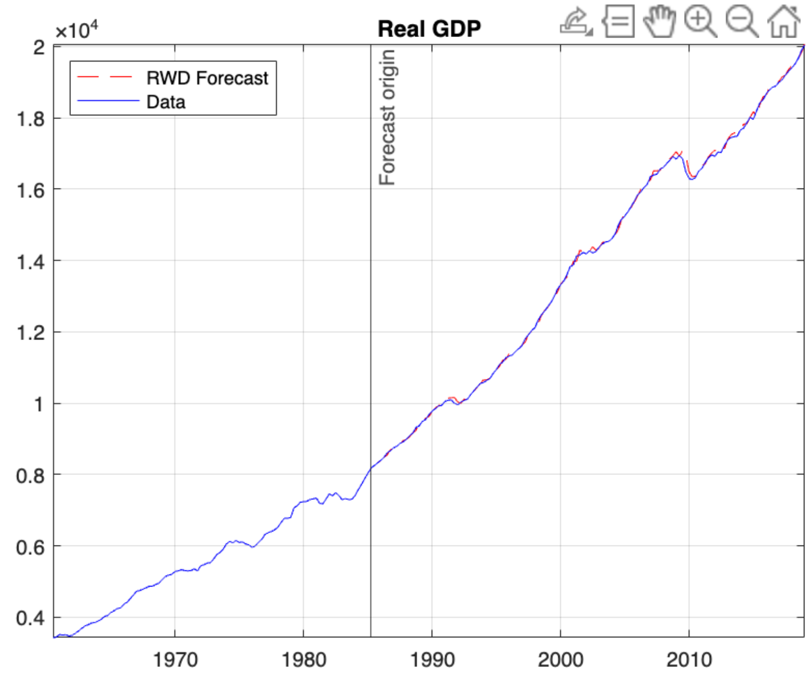
$$y_t = \ c + y_{t-1} + \epsilon_t \quad \text{with } \ \epsilon_t \sim \text{iid } N(0, \sigma^2)$$

(is equivalent to specify an arima model for returns: ARIMA(1,1,0))

Forecast

$$\hat{y}_{T+h|T} = E(y_{T+h}|T) = \mathrm{h}\hat{c} + y_T \quad \text{with } h = 1$$

Real GDP Vs RWD forecast



$$RMSE_{RWD} = 82.0144$$

ROOT MEAN SQUARED ERROR (RMSE)

Root Mean Squared Error (RMSE) - formula

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum\left(\text{variable}_{t,observed} - \text{variable}_{t,forecasted}\right)^2}$$

	RW	AR(4)	VAR(4)	VAR(p)	AR(4)-X	Bonus: RWD
RMSE	120.06	80.38	92.91	87.98	83.95	82.01

The most powerful model is that one with the minor root mean squared error, that, according to the previous predictions, is the AR(4) model.