UNIVERSITÀ
DEGLI STUDI
DI MILANO

Time Series Empirical Project

Program: Data Science for Economics

2023

Authors:	*ID:*
Giordano Vitale	14310A
Pietro Padovese	12356A

Course Coordinator:

Prof. Andrea Bastianin

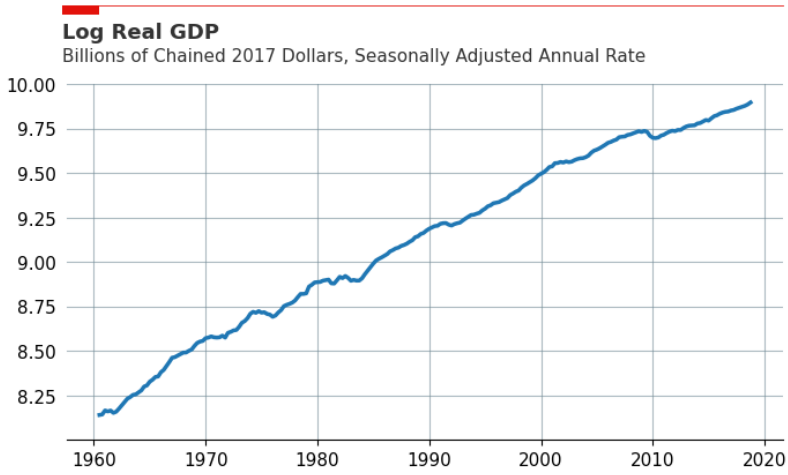
15th December 2023

Contents

Section 1

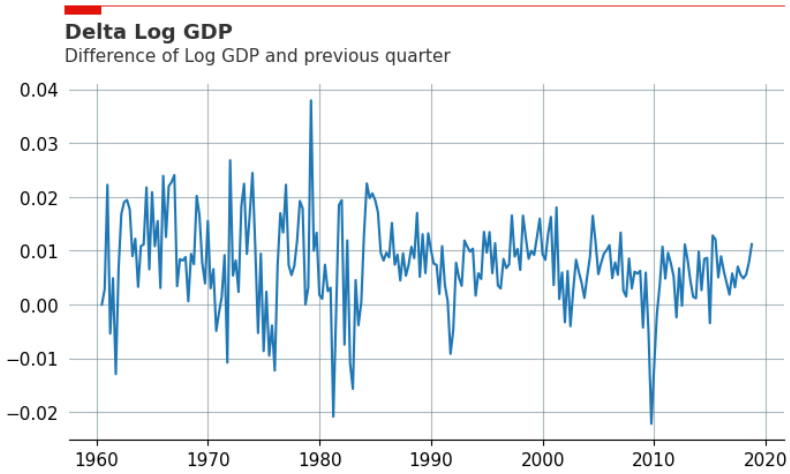
This section contains the plots for the most important series in the empirical analyses. All the series are expressed in quarterly data and they span from 1960-Q2 until 2018-Q3.

- **Plot** $y_t = log(RealGDP_t)$



This series exhibits an upward trend over its entire period. A few troughs can be noticed, especially around 1976, 1984, and 2009, suggesting periods of economic contraction or recession. As a consequence of the *log-transformation*, the y-scale becomes less interpretable, yet the overall trend is still significantly observable.

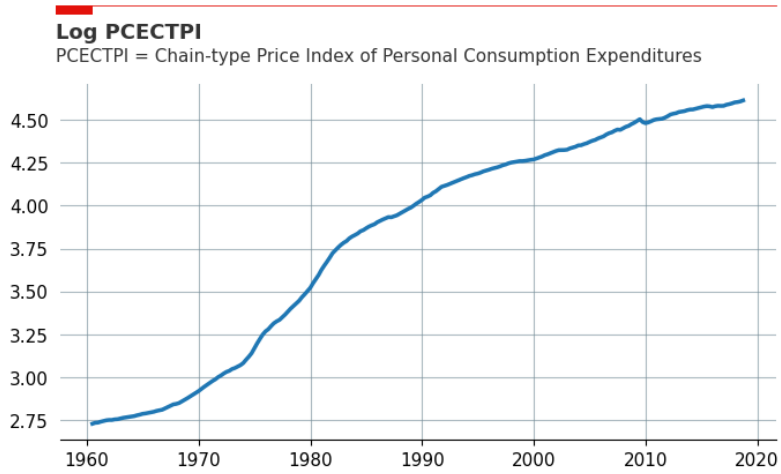
- **Plot** $\Delta y_t = \log(RealGDP_t) - \log(RealGDP_{t-1})$



This series represents the first difference of the aforementioned *log-transformed* Real GDP and it doesn't exhibit trends, not cycles, or seasonality. This aligns with the theoretical expectations

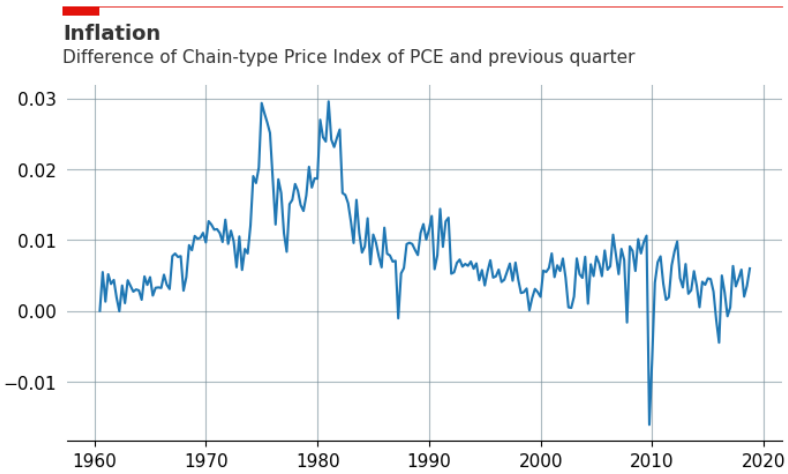
for a first difference. The series fluctuates around its mean (0.07): a positive mean implies that, over the observed period, the economy has, on average, experienced overall positive proportional changes in GDP.

- **Plot** $\log(PCECTPI_t)$



The consistent upward trend in $\log(PCECTPI)$ indicates that, on average, the overall level of prices for goods and services in the consumption basket has risen over time. PCECTPI is often used by economists to have a general idea about how prices move over time. An increasing PCETCPI means that, on average, consumers are experiencing higher prices for the goods and services they typically purchase.

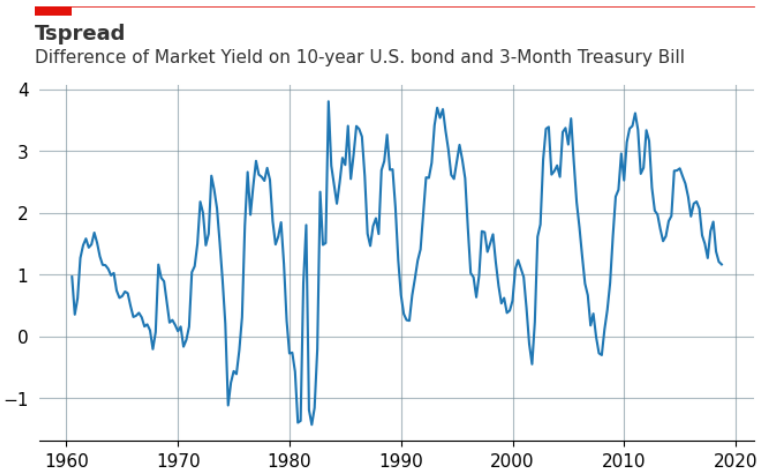
- **Plot** $\pi_t = \log(PCECTPI_t) - \log(PCECTPI_{t-1})$



This series captures the approximate percentage change in inflation growth. In this series it is possible to spot cyclical patterns: in fact, periods of decreasing inflation growth are combined

with periods of increasing inflation growth. It's worth underlying a specific outlier around 2009, which is coherent with the deep recession period occurred at that time. Except for this negative outlier, in the period after 1990 there have been negligible percentage changes in inflation growth, as one can see from the low *volatility cluster* in the specified period. This pattern is consistent with an era of more stability, in terms of prices, with respect to previous decades.

- **Plot** $Tspead_t = GS10_t - TB3MS_t$

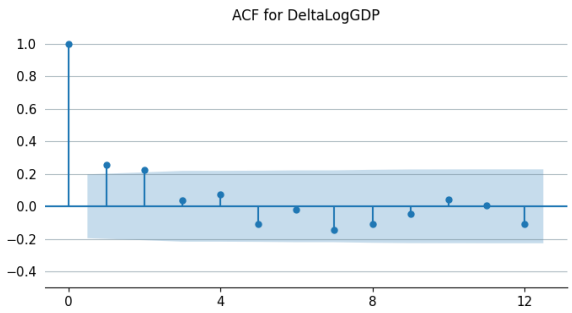


Tspread stands for *Term Spread* and it quantifies the difference in interest rates between long-term (10-year) and short-term (3-month) Treasury securities. In a "normal", "stable" economic period, long-term interest rates are higher than short-term rates, resulting in a positive Term Spread. This scenario is often associated with expectations of economic expansion. Conversely, an inverted yield curve occurs when short-term rates are higher than long-term rates, resulting in a negative Term Spread, which is often considered a potential precursor to an economic downturn. In this series we can notice several cycles: if the Tspread moves from negative or lower values to positive values, this is associated with positive expansion cycles. Vice versa, if Tspread moves from high values to lower or negative, this is associated with bad economic periods. In addition, positive values of Tspread are more often observed than negative values, coherently with the upward increase in GDP over time mentioned in the previous comments.

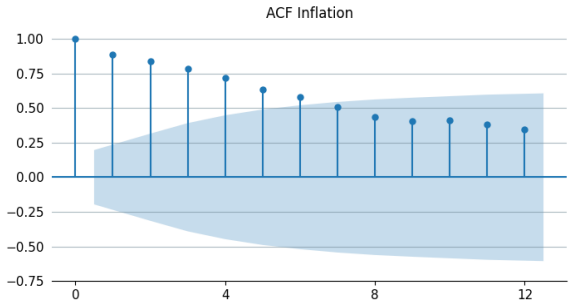
Section 2

In this section, we first plot ACFs for the following series: $\Delta y_t, \pi_t, Tspread_t$. In all cases, the number of lags=12 is selected for the plots. Secondly, we use the *AIC* criterion to select the lag order for a VAR(p) model, which is in turn fitted and whose residuals will be graphically and qualitatively analyzed through ACF and PACF.

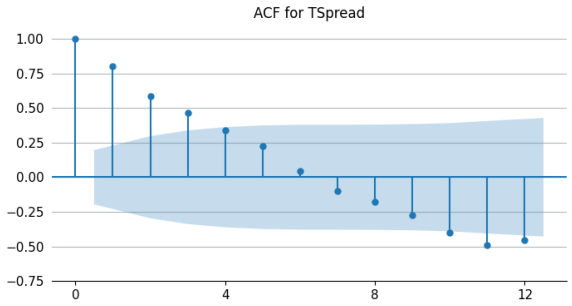
- **Sample AutoCorrelation Functions (ACFs)**



DeltaLogGDP doesn't show a significant degree of similarity with its lagged values. The relationship between present and past values is weak, and it is significant only for lag 2 and 3, even though the coefficients are quantitatively low. This pattern is coherent with a series with no cycles, trends, or seasonality, such as *DeltaLogGDP*.



π_t shows a significant and quantitatively high - though gradually decreasing - degree of similarity with its first 6 lagged values. The relationship between present and past values is stronger than in the previous case, meaning that there is a significant positive autocorrelation, as all the coefficients are bigger than zero. From lag 7, the coefficients become non-statistically significant. This pattern is coherent with a mean-converging series.

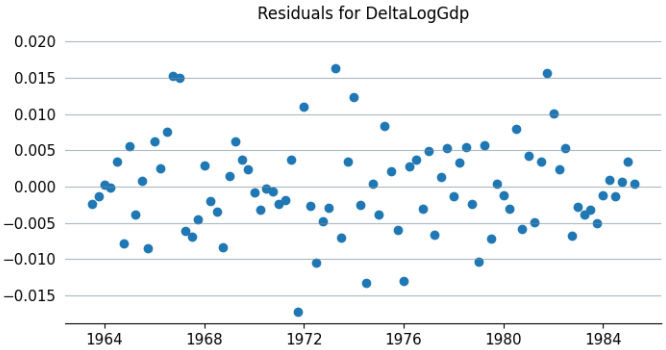


The first three autocorrelation coefficients for $Tspread$ are strongly positive - though gradually decreasing - meaning that the relationship between present and past values is statistically different from zero only within one-year-window. It is also interesting to underline that three negative autocorrelation coefficients are statistically different from zero, suggesting a negative relationship between present values and past values for almost three years before. This pattern is coherent with a series that follows cyclical fluctuations.

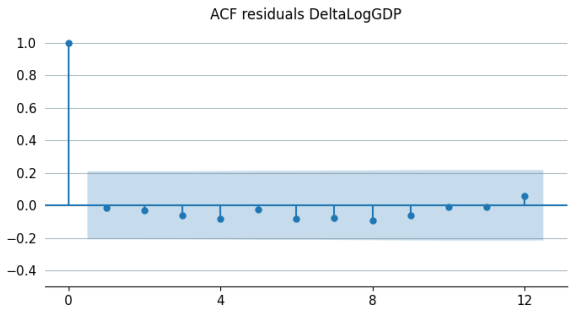
- **Model Selection with AIC & residuals analysis**

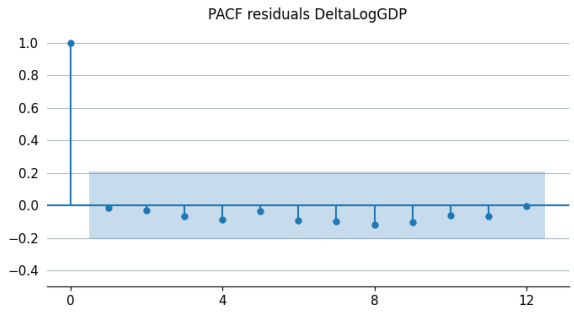
Using AIC criterion, the number of selected lags is p=12. We then proceed by fitting a *VAR(12)* model on the first window of 100 observations and obtain the corresponding residuals.

Residuals analysis for *DeltaLogGDP*



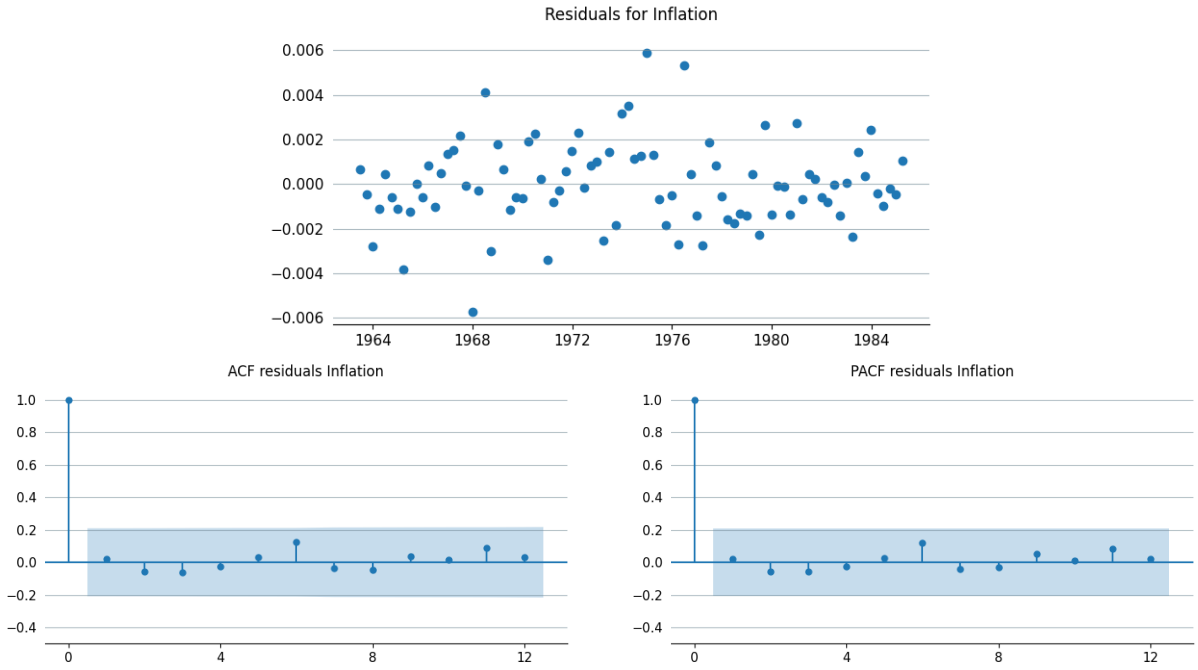
The scatter plot of residuals for *DeltaLogGDP* reveals a lack of discernible patterns, indicating that the VAR(12) model residuals don't suffer from autocorrelation. The absence of outliers and a distribution around 0 imply that the model residuals exhibit no systematic bias. The homoscedasticity-like distribution of the data points suggests that the model's coefficients and their confidence interval can be considered statistically reliable. In summary, the residual diagnostics have yielded satisfying results in terms of model accuracy, statistical significance, and absence of autocorrelation.





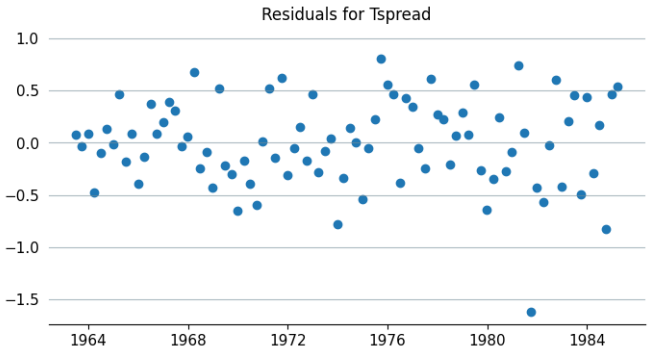
As regards ACF and PACF of the residuals for *DeltaLogGDP*, they both have non-significant coefficients, ruling out serial correlation, as we mentioned above. A further element backing the adequacy of the model is represented by the non-significant partial autocorrelation coefficients, which exhibit the same behavior of the ACF coefficients.

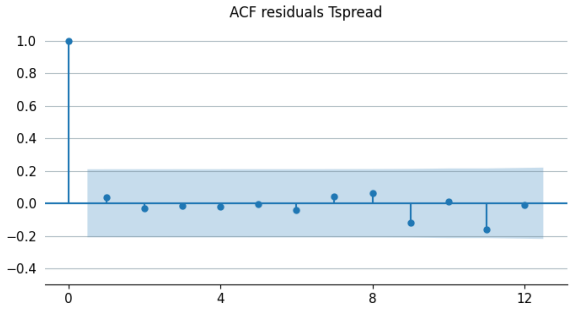
Residuals analysis for *Inflation*

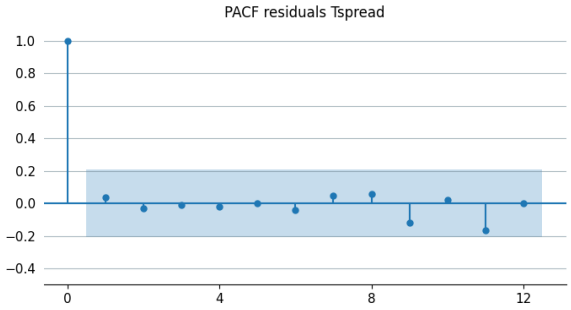


The scatter plot, as well as the ACF and PACF for *Inflation*, exhibit the same properties discussed for *DeltaLogGDP*, and this is a positive result confirming the goodness of the model, its statistical significance, and reliability in the coefficients.

Residuals analysis for *Tspread*



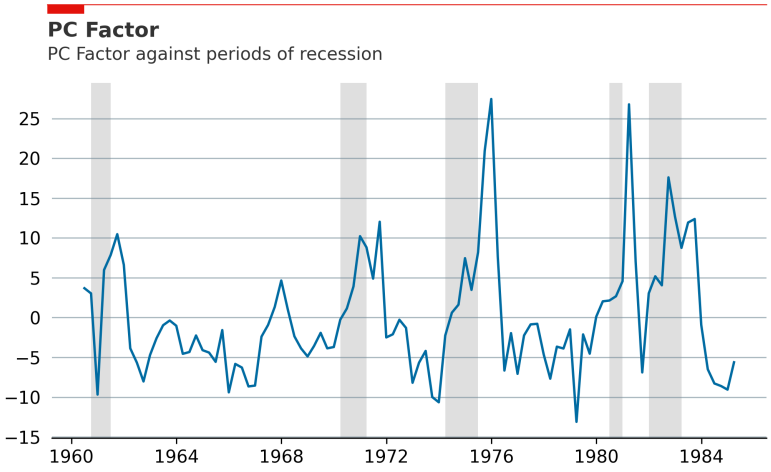




Also in this case, analyzing the residuals for *Tspread* we can easily notice that it is distributed around 0, without clear and definitive patterns. One main difference this scatter plot presents with respect to the previous ones is the presence of one outlier at the time 1982. Except for this data point, the variability is limited and bounded between -1 and 1. The ACF and PACF yield the same conclusion derived for *DeltaLogGDP* and *Inflation*, as in all lags the coefficients are statistically non-different from zero.

Section 3

Plotting PC Factor



In order to understand what this plot may tell us, it is useful to start from the framework of the dynamic factor model. It operates on the premise that a small number of common factors drive the co-movements of a much larger number of time series variables. These factors are assumed to be unobserved but are estimated through the principal components of the observed series. For this analysis the principal component has been computed using more than 250 economic series, each potentially containing information that characterizes the state of the economy.

In this particular example, only the first principal component has been used, as it exhibits the highest explanatory power, making it the key driver in capturing the main behavior of the series. Notably, the principal component factor doesn't display any significant trend or seasonality, indicating that the economic series used are generally stable within the considered timeframe.

What seems to emerge are irregular cycles characterized by medium to long periods of relatively small jumps in the series, followed by rapid and substantial upward shifts. To better understand this behavior, periods of recession (Fred data) have been highlighted in gray. They align notably well with instances when the principal component factor sharply rises[1]. This suggests that the

[1]From the graph, it actually seems that peaks in the principal factor occur with a certain lag compared to recession periods. However, the data used in the analysis (taken from the Excel file) appears to be shifted by one

co-movement captured by the first principal component reflects the behaviour of the business cycle, which is characterized by periods of expansion followed by periods of recession occurring simultaneously in many economic activities.

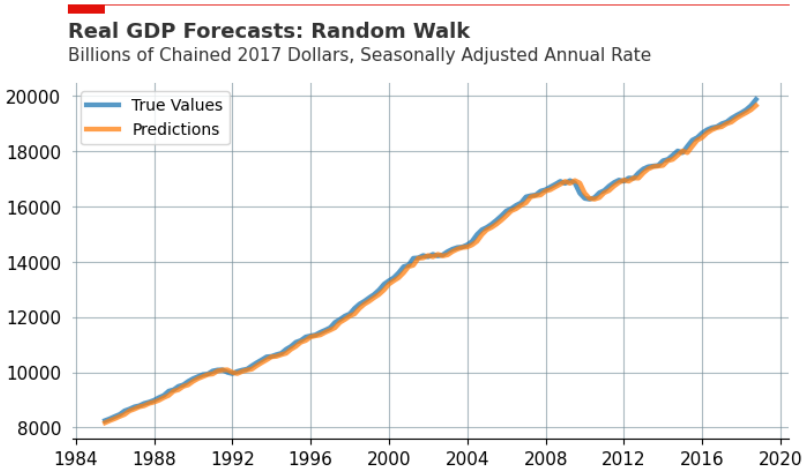
year compared to the original FRED data, which would explain the misalignment between the principal factor and recession periods. Nevertheless, we chose to use the data from the Excel file instead of taking them from FRED to ensure consistent results.

Section 4

In this section all the estimated models are shown and compared.

- **Random Walk**

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

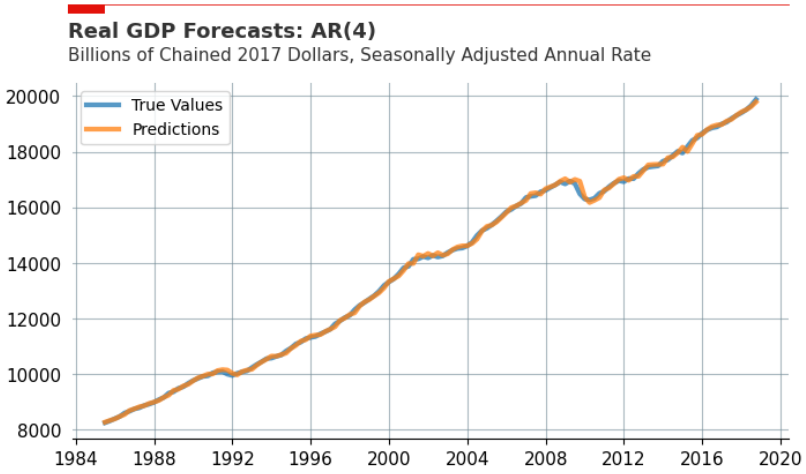


- **AR(4)**

$$\Delta y_t = \alpha + \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \rho_3 \Delta y_{t-3} + \rho_4 \Delta y_{t-4} + u_t \qquad u_t \sim WN(0, \sigma^2)$$

α = intercept

ρ_i = parameter



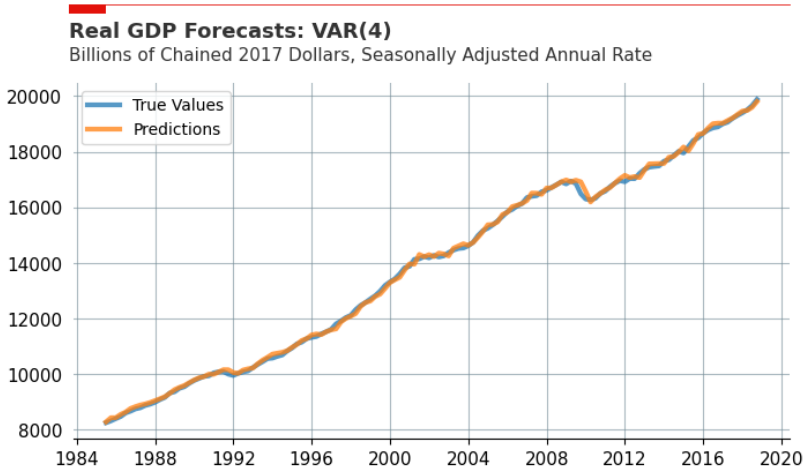
- **VAR(4)**

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + A_3 Y_{t-3} + A_4 Y_{t-4} + U_t$$

$$Y_t = \begin{pmatrix} \Delta y_t \\ \pi_t \\ Tspread_t \end{pmatrix} \qquad A_i = \begin{pmatrix} a_{i,11} & a_{i,12} & a_{i,13} \\ a_{i,21} & a_{i,22} & a_{i,23} \\ a_{i,31} & a_{i,32} & a_{i,33} \end{pmatrix} \qquad U_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{pmatrix}$$

The coefficient $a_{t,mn}$ stored in matrix A_i represents the effect that the *m-th* variable at lag *i* has on the *n-th* variable at time *t*.

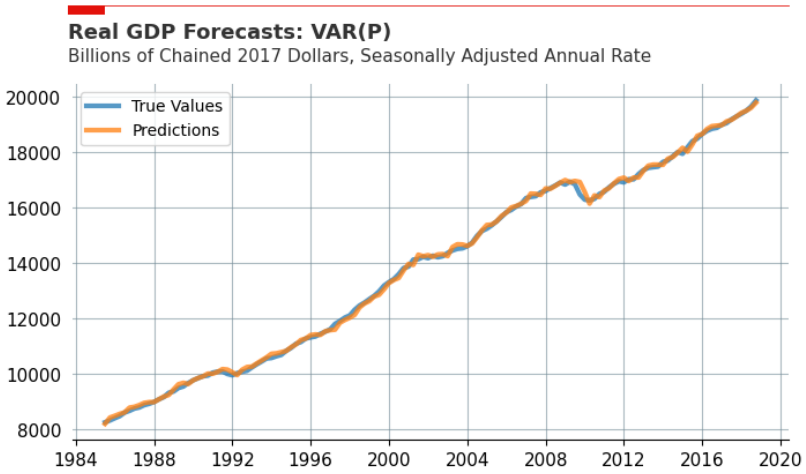
u_i, t represent White Noise processes with mean = 0 and variance = σ^2.



- **VAR(p)**

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Yt - p + U_t$$

The specification of the model is identical to the one used for the VAR(4) model, with the only exception that in this case the lag order p is selected every time we produce a forecast as the one that returns the lowest AIC.



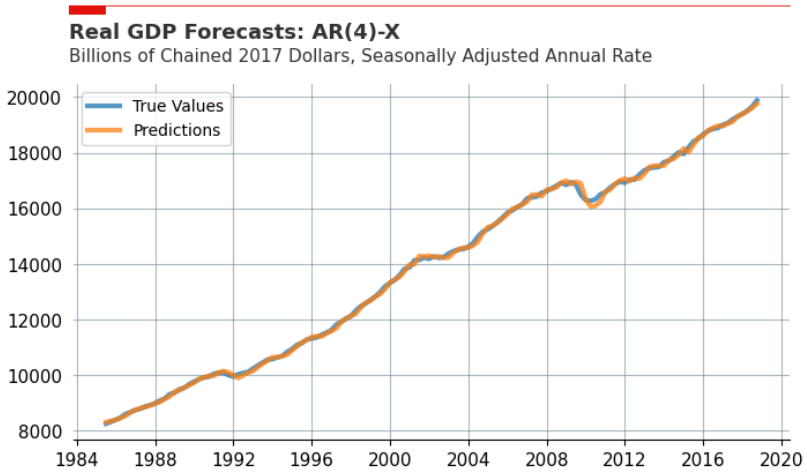
- **AR(4)-X**

$$\Delta y_t = \alpha + \rho_1 \Delta y_{t-1} + \rho_2 \Delta y_{t-2} + \rho_3 \Delta y_{t-3} + \rho_4 \Delta y_{t-4} + c\hat{F}_{1,t-1}$$

α = intercept

ρ_i, c = parameters

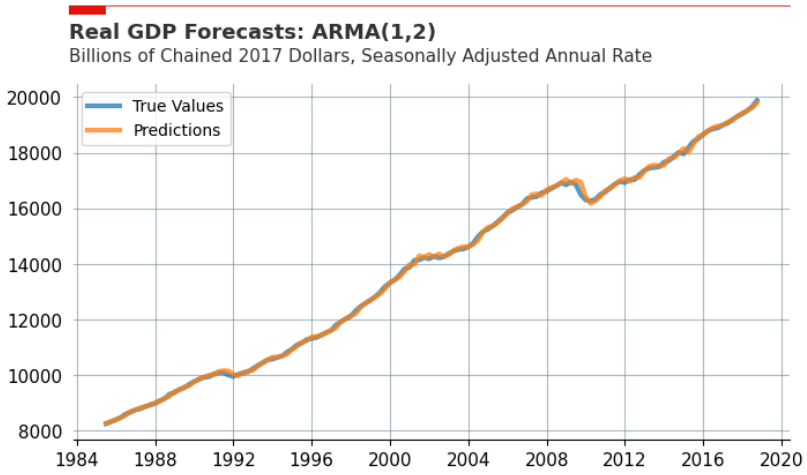
$\hat{F}_1$ = first principal component



- **ADDITIONAL MODEL: ARMA(1,2)**

$$\Delta y_t = \alpha + \rho_1 \Delta y_{t-1} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \qquad u_t \sim WN(0, \sigma^2)$$

Analyzing the plots of ACF and PACF for Δy_t, in both there is a cut-off after few lags, suggesting that an appropriate model for forecasting is one containing both the autoregressive and moving average parts. In detail, the q and p orders were selected by examining the numbers of significant lags in the ACF and PACF plots, respectively.



RMSE

Model	RW	AR(4)	VAR(4)	VAR(p)	AR(4)-X	ARMA(1,2)
RMSE	119.44	79.98	93.38	97.83	83.75	79.45

The table reported above displays the root mean squared errors for the different models considered. Following the order in which they are reported, the Random Walk model reports the highest RMSE, and is therefore the model with the worst predictive performance. This is probably due to the over-simplicity of this model, as it uses just the last available observation as the forecast.

The AR(4) model turns out to be the second best predictive model, close behind the first. This good result means that the changes in GDP in the previous four quarters are a good indicator of the trend it will have in the immediate future.

The VAR(4) model, which tries to simultaneously predict the values of Δy_t, inflation and spread, is less effective than the AR(4) model. In this, the difficulty of modeling multiple variables simultaneously is not sufficiently offset by the advantage of being able to use information derived from multiple economic series.

When the number of lags to be included in the VAR model is chosen through information criteria, the result gets even worse. When analyzing the number of lags chosen in the different iterations, it appears that this number often seems to be excessively high, a factor that can cause overfitting problems as the model adheres excessively to the training data. This result could be caused by the information criteria used, namely the AIC, which uses a penalty for the number of predictors in the model lower than Bayes' Information Criteria.

The AR(4) model using the principal factor as an exogenous variable, although it performs better than the VAR models, fails to perform as well as the original AR(4). This indicates that the information contained in the principal factor does not help improve forecasts. Based on what was said in section 3, one possible explanation is that the information contained in the PC "overlaps" with that contained in the lagged variable of Δy_t. If so, the addition of the principal factor would only increase the complexity of the model, without adding explanatory power.

The proposed ARMA(1,2) model performs better than all the others, although only slightly compared to the AR(4). This result may be due to the fact that by introducing a component of moving averages, represented by past error terms, it is possible to capture the impacts of short-term economic shocks that are not entirely explained by autoregressive components alone.