Predict the business cycle: relationships between real and nominal variables

Final Project of Time-series Analysis of Economic and Financial Data

Group members: Crespi (3018540), Molaro (3034820), Mulazzani (3026084), Sirugo (3013547)

Introduction

The aim of this project is to study the possible relationships between real and nominal economic variables in order to be able to describe and make predictions on the future movements of the business cycle. In other words, we describe the business cycle considering the Italian GDP growth rate and the unemployment rate, then the business cycle influences the decisions of monetary policy. Thus, these changes in the nominal variables have an impact on the business cycle itself and so on the real economy. In order to achieve this goal we need to correctly define the most suitable statistical model to use for the analysis. The following sections will cover respectively: a brief description of our dataset, then we will apply an Hidden Markov Model on the real variables using three states and to overcome the assumption of abrupt change points we will implement a Dynamic Linear Model. In the final part of the project we try to answer our initial question regarding the relationship between the GDP growth rate and the monetary variables, that is BOT and BTP interest rates and inflation. This will be done firstly with a classical regression model and then we will introduce time dependence in the parameters of the regression.

Data presentation

For this analysis we are going to consider a dataset which combines different time series such as the Italian GDP growth rate, unemployment rate (seasonal adjusted), three-month and ten-year government bonds' interest rates (BOT and BTP respectively) and the inflation rate (using the CPI). All these variables are measured quarterely from 1996 to 2019 for Italy. In particular, different sources have been used: OECD for GDP, BTP and BOT interest rates, ISTAT for the unemployment rate and FRED for CPI. In the following table some basic summaries have been collected:

Variable	# observations	Min	Max	Mean	Std. Deviation
Unemployment Rate (%)	96	5.949	12.739	9.633	1.897
GDP Growth Rate (%)	96	-2.795	1.696	0.134	0.675
BTP Interest Rate (%)	96	1.213	10.545	4.306	1.757
BOT Interest Rate (%)	96	-0.403	9.950	2.391	2.443
Inflation Rate (%)	96	-0.390	5.176	1.838	1.087

We plot all the variables in a separate graph with the exception of the government bonds rates which are plotted together; the BOT Interest rate in red and the BTP one in blue respectively. They clearly show an opposite trend between the GDP growth rate and the unemployment and also the return of bonds tend to increase during a recession since the risk premium required by investors increases.

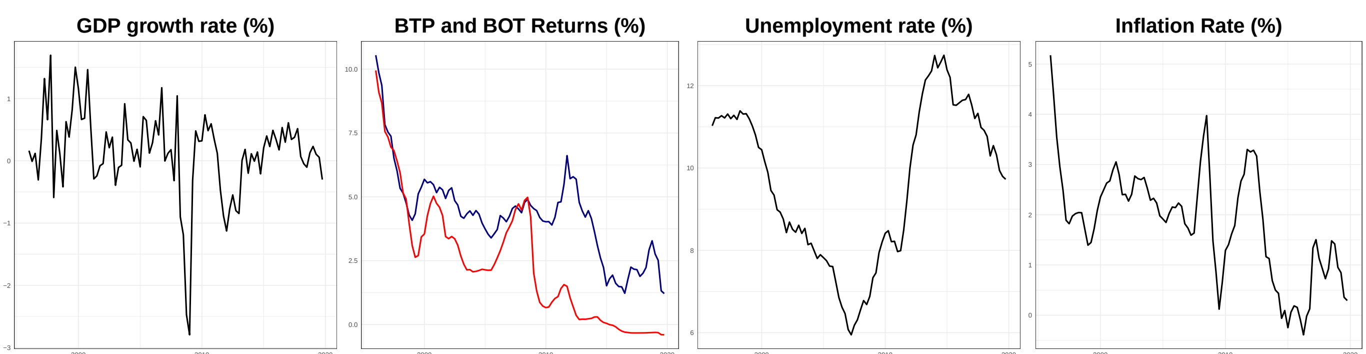


Figure 1. Time Series of nominal and real economic variables

Hidden Markov Model

In this section we implement a Gaussian Hidden Markov Model for the Italian GDP growth rate and the unemployment rate to check if the results on the first variable are consistent with the second one. The reasons behind the choice of this model, as mentioned above, are the seemingly non-stationarity of the series and the presence of change points, particularly in the GDP growth rate. In particular, we can recall the 2007-08 crisis and the sovereign debt crisis in 2012 in Europe. Thus, we decided to implement a three-state model representing the three classical states of the economy: recession, stability and expansion. The model specification follows the common three steps: identification (we have already defined the states), estimation and decoding. The model is the following:

$$\begin{cases} Y_t = \mu_1 + \epsilon_t \quad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma_1^2) \quad if \quad S_t = 1 \\ Y_t = \mu_2 + \epsilon_t \quad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma_2^2) \quad if \quad S_t = 2 \\ Y_t = \mu_3 + \epsilon_t \quad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma_3^2) \quad if \quad S_t = 3 \end{cases} \tag{1}$$

The following tables contain the estimated transition probabilities and the estimated emission parameters for the HMM constructed for the GDP growth rate and the unemployment rate. From the first table we can notice that the estimated transition probabilities are higher in the main diagonal suggesting that the latent process is more likely to remain in the same state, however this is sligthly less true for the GDP growth rate.

	HMM on GDP						HMM on Unemployment					
t-1/t	State 1	S.E.	State 2	S.E.	State 3	S.E.	State 1	S.E.	State 2	S.E.	State 3	S.E.
State 1	0.925	0.057	0.009	0.031	0.066	0.072	0.926	0.072	0.000	NA	0.074	0.072
State 2	0.189	0.121	0.811	0.121	0.000	NA	0.000	NA	0.980	0.019	0.020	0.019
State 3	0.068	0.103	0.064	0.072	0.868	0.129	0.033	0.033	0.033	0.033	0.934	0.046

Besides this, the estimated parameters are interesting because they point out the expected value of Y_t (observed process) conditional to a certain latent state. Namely, the expected value of the GDP growth rate conditional to the state two is equal to -1.076% indicating a recession. While, for the unemployment rate this expected value, conditional to state two, is equal to 11.184% again underlying the consequences of a recession.

	HMM on GDP				HMM on Unemployment			
States	Intercept	S.E.	St. deviation	S.E.	Intercept	S.E.	St. deviation	S.E.
State 1	0.187	0.056	0.284	0.035	6.653	0.177	0.442	0.127
State 2	-1.076	0.264	0.770	0.177	11.184	0.120	0.799	0.089
State 3	0.545	0.165	0.591	0.107	8.303	0.109	0.457	0.087

Finally, we plot the results obtained. They show the original series together with the estimated sequence of states and the 95% credible interval. The two graphs below are interesting since they confirm our initial observations about the behavior of these two time series. In fact, the model underlines both the two crisis in the GDP graph and the same pattern is highlighted for the unemployment rate with a lag with respect to the GDP growth rate.

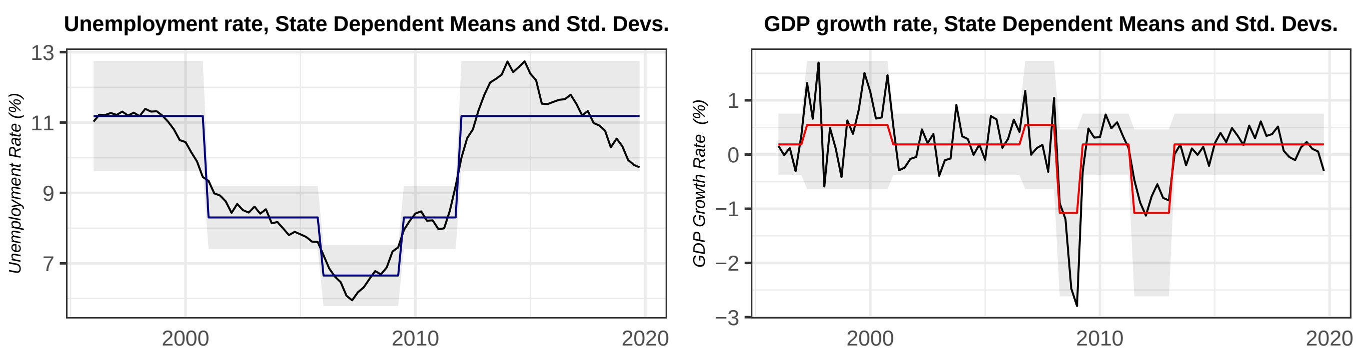


Figure 2. Plot of estimated results of HMM for GDP growth rate and Unemployment rate

Dynamic Linear Model

In this section we are going to implement a Dynamic Linear Model. In fact, since the assumptions of state space models do not require stationarity of the state process, the local level model (which is a state space model) may work well in our setting. From now on the analysis will focus on the GDP growth rate while the unemployment rate is no longer useful since it was just used as a check in the HMM paragraph. Thus, in order to overcome the assumption of abrupt changes in the hidden state, we assume it to be continuous over time to be able to implement a dynamic linear model on GDP growth rate. The major shortcoming of the model we have just studied is that the state space is discrete. Trends and relations between economic variables are poorly explained under a discrete time probabilistic approach. We now employ a random walk plus noise model to analyse the GDP growth rate. Looking at the plot of the GDP growth rate we see that it does not exhibit any evident seasonality or trend component. Thus, we believe that the random walk plus noise can be a good model in our setting. The model is characterized by the following observation and state equations:

$$\begin{cases} Y_t = \mu_t + v_t & v_t \sim N(0, V) \\ \mu_t = \mu_{t-1} + w_t & w_t \sim N(0, W) \end{cases} \tag{2}$$

where (v_t) and (w_t) are independent. Moreover, recall that this is a so-called steady model since the forecast function is constant, that is $f_t(k) = E(Y_{t+k} \mid y_{1:t}) = m_t$. Thus, this is the main disadvantage of this model since it is obvious that we cannot expect a constant value of the GDP growth rate. More formally, in order to implement a DLM as described above we have firstly computed the MLE for the unknown parameters, that is V and W. In particular, the results obtained are: $\hat{V} = 0.1245$ with an asymptotic standard error of 0.0478, and $\hat{W} = 0.1646$ with asympotic standard error of 0.0639.

Once we have obtained these estimates then we have used them to build the random walk plus noise model. After this step, we have implemented the Kalman smoother to recover the past history of the latent state process. The choice of using the Kalman Smoother and not the Kalman Filter is related to the fact that we are dealing with "batch" data and not with "streaming" data.

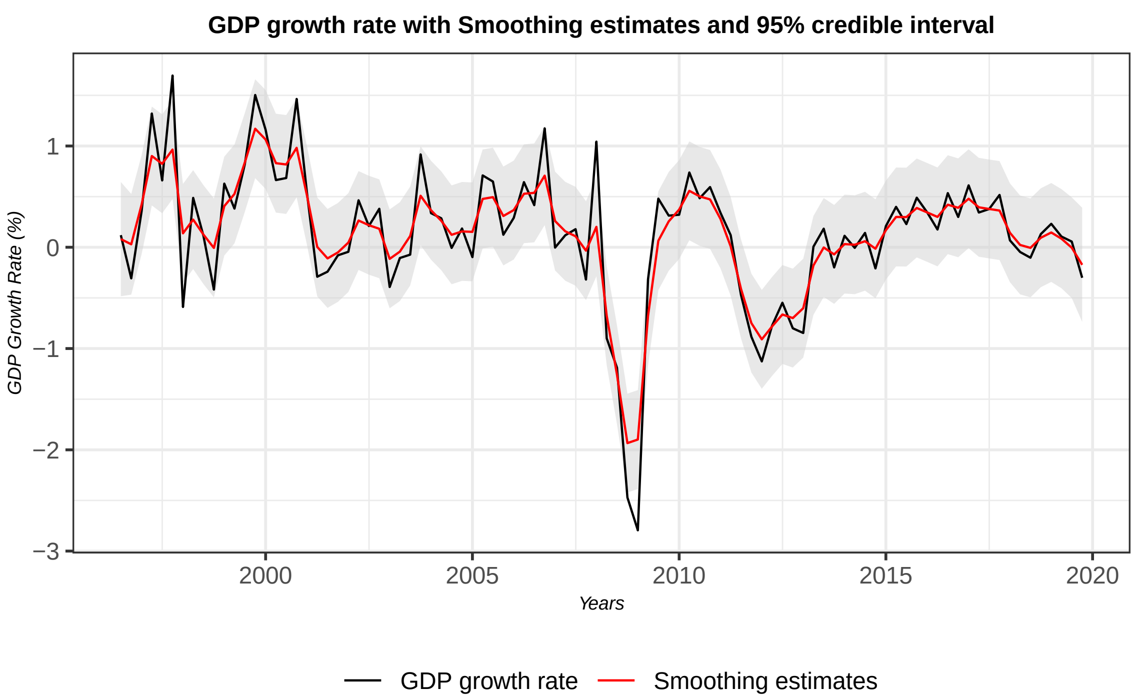


Figure 3. Plot of smoothing estimates

Linear Regression Model

Proceeding in our analysis, we introduce monetary variables, which we believe can help us explain the path of GDP growth rate over time. To do so we implement a static linear regression model using three regressors and including also the intercept. In particular, the regressors included are: BOT and BTP interest rates and the CPI rate. The dependent variable described is the GDP growth rate. An important assumption of our model is that GDP growth rate is related to the monetary variables with a lag of two periods, that is two quarters, since we believe that the trasmission mechanism of monetary policies into the real economy takes time to manifest its effects. For example, the values of the nominal variables registered in the first quarter of 1996 are used to describe the GDP growth rate in the third quarter of 1996 and so on. In other words, the static linear regression we are going to implement is:

$$GDP_t = \alpha + \beta_1 BOT_{t-2} + \beta_2 BTP_{t-2} + \beta_3 CPI_{t-2} + \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{3}$$

The following table shows the main summaries of the regression model (3):

Regressor	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	0.220313	0.209582	1.051202	0.295980	
BOT Interest Rate (%)	0.052978	0.045467	1.165197	0.247018	
BTP Interest Rate (%)	0.175876	0.078334	2.245211	0.027204	* (at 0.05 level)
Inflation Rate (%)	-0.525839	0.096086	-5.472599	0.000000	*** (at 0.001 level)

This summary table underlines that the BOT interest rate is not a statistical significant regressor and so we try to drop it from the analysis. This decision is also supported by the fact that considering a lagged GDP growth rate of two quarters, makes the interest rate of a three-month bond a bit meaningless. Thus, we update our model to:

$$GDP_t = \alpha + \beta_1 BTP_{t-2} + \beta_2 CPI_{t-2} + \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{4}$$

The summary table of this new regression (4) is:

Regressor	Estimate	Std. Error	t value	Pr(>|t|)	Significance
Intercept	0.080694	0.172288	0.468370	0.640640	
BTP Interest Rate (%)	0.234533	0.060134	3.900142	0.000184	*** (at 0.001 level)
Inflation Rate (%)	-0.518850	0.096087	-5.399813	0.000001	*** (at 0.001 level)

We can now notice that all the regressors are highly significant. Using the Information Criteria to compare the two models, we have observed that the AIC score (175.277) of the model with two regressors (4) is better than the AIC score (175.869) of the model with three regressors (3). This is in line with the mostly intuitive macroeconomic concepts, since the GDP of a country is explained by a multitude of variables.

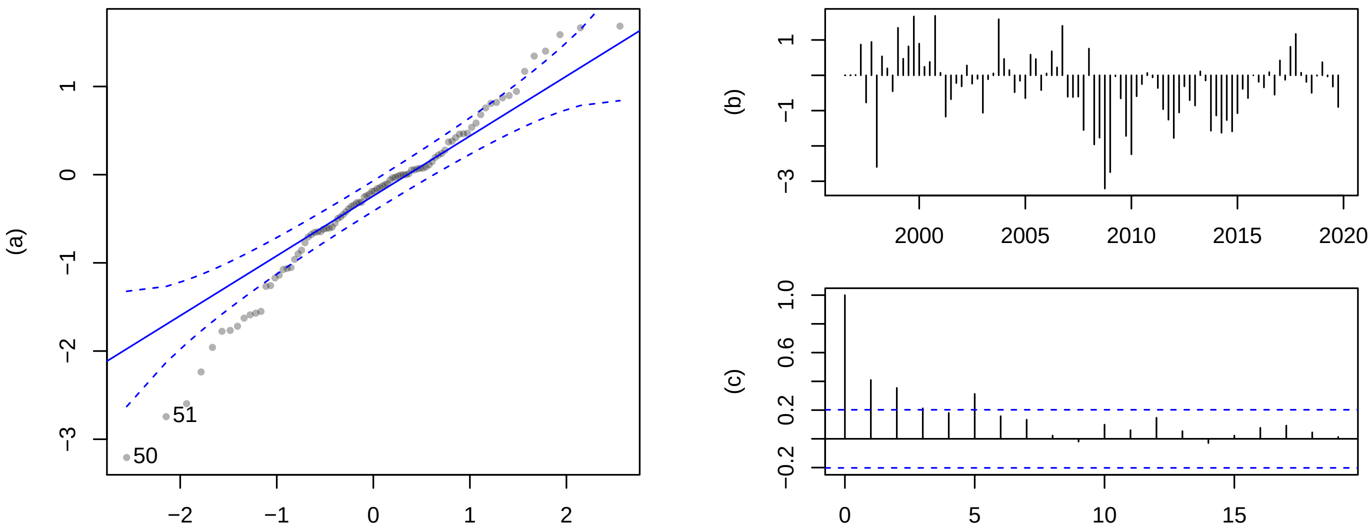
Bayesian Linear Regression

Now we model a bayesian linear regression considering the following generic model:

$$\begin{cases} Y_t = X_t' \, \theta_t + v_t & v_t \overset{i.i.d}{\sim} N(0, \sigma_t^2) \\ \theta_t = \theta_{t-1} + w_t & w_t \overset{ind}{\sim} N_3(0, W_t) \end{cases} \tag{5}$$

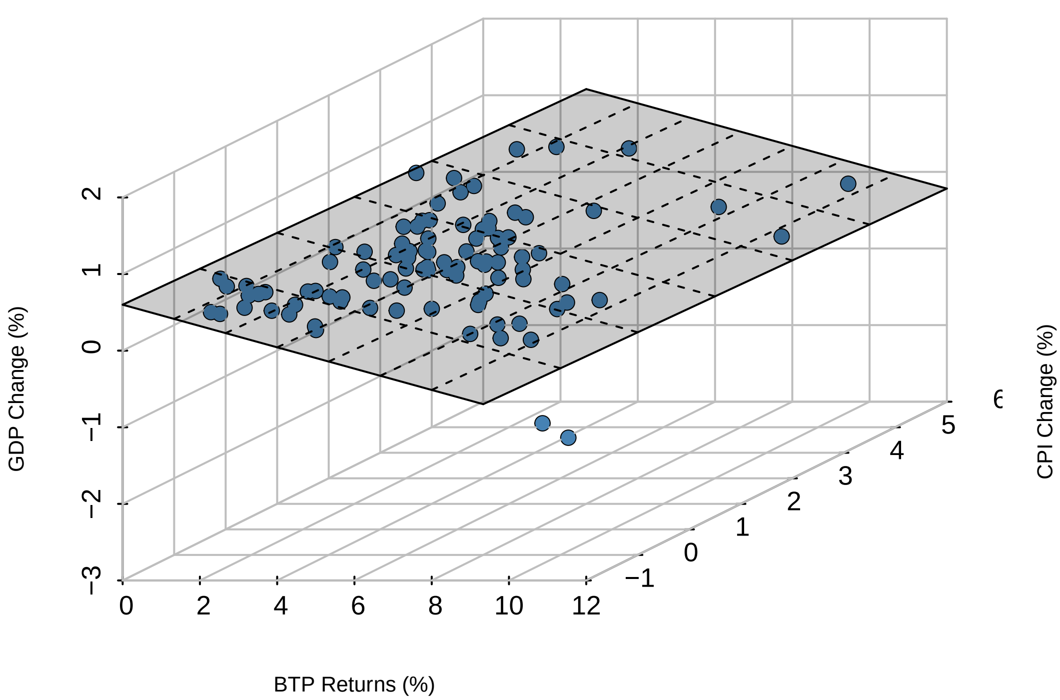
where θ_t is a vector of the parameters α, β_1 and β_2. In this specific case we will set W_t equals to zero since we want constant parameters and thus we simply model the above specified classical static regression (4) as a DLM. In this context we will have that the state equation is no longer useful in describing the model, since our θ will be constant over time. From a practical point of view, we firstly estimate the unknown parameter by MLE computing also its asymptotic standard error, that is $\hat{\sigma_t}^2 = 0.3585$ with aymptotic standard error of 0.05314. Then we use this estimate to run the DLM obtaining the coefficients of the regression as follows: $\alpha = 0.0807$, $\beta_1 = 0.2345$ and $\beta_2 = -0.5189$

Moreover, to check the goodness of the model we implement some basic strategies aimed at testing the normal distribution of the residuals and the absence of correlation among them. Thus, we present the Q-Q Plot, the correlogram of the autocorrelation function of the residuals and the plot of the residuals. As we can see the Q-Q Plot looks quite good despite showing some mild skewness and we can notice the outliers corresponding to the second and third quarters of 2008 during the financial crisis. The following plots are: (a) Q-Q Plot, (b) Standardized residuals, (c) ACF.



Note that we also implement the Shapiro-Wilk test in order to check the normality also through a statistical test, and we obtain a p-value of 0.07103. This means that we cannot reject the null hypothesis, that is the residuals are normally distributed, at the 5% statistical level. As regards the correlogram we can notice that there is some autocorrelation between residuals and thus an autoregressive model could solve the problem and perform more efficiently.

Finally, to have a graphical representation of the correlation between the changes in the monetary variables (BTP returns and inflation change) and the change in GDP, we plot the following 3D scatter plot. Moreover, to highlight the marginal effects we plot them too.



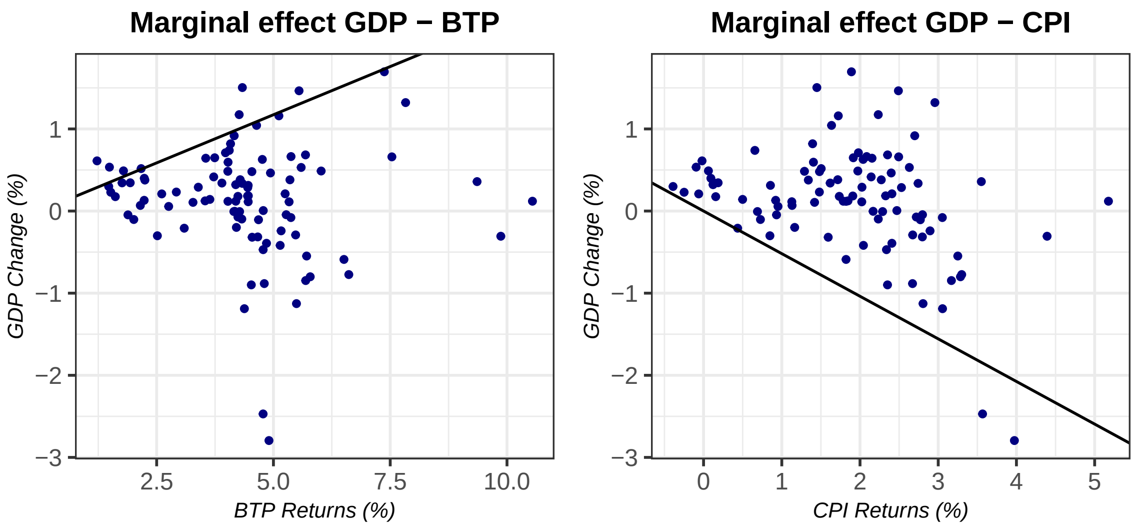


Figure 4. Plots of marginal effects.

From the above scatter plot we can observe how data are distributed near the regression plane, except for the two outliers registered during the period of the finacial crisis in 2007-08.

Dynamic Regression Model

As a final step we decided to improve the above regression model letting the parameters change over time. We believe that during different times the impact of monetary variables on GDP growth rate could have changed, in particular we feel that this may have been a relevant change during the financial crisis of 2007-08, when we expect our coefficients to decrease in magnitude. In other words, the intensity of the relationship between variables is changing over time. Thus, we implement the same regression as above, the one with two indipendent variables, namely BTP and CPI, as a dynamic regression model.

$$GDP_t = \alpha_t \; + \; \beta_{1t} \; BTP_{t-2} \; + \; \beta_{2t} \; CPI_{t-2} \; + \; \epsilon_t, \qquad \epsilon_t \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{6}$$

This regression has been implemented using a Dynamic Linear Model, the specification is the same of the one used in the previous paragraph for the static regression (4), but now the matrix W_t of the errors of the state equation is not a null matrix. This assumption corresponds to introducing time-varying coefficients. Moreover, in this simplified model we assume the covariates of this matrix to be all zeros, this could be considered as a limit of the model since it is natural to think that the above regressors are related.

Again, the implementation of this DLM has followed the usual steps: the computation of MLEs and their asymptotic standard errors, and then these estimates have been used to compute the coefficients through the Kalman smoother. As regards the results of the Maximum Likelihood Estimates, we obtain:

$$\begin{aligned}
&\hat{\sigma}^2_v = 0.10367 && Asymp.\ Std.\ error = 0.0353 \\
&\hat{\sigma}^2_{w\alpha} = 0 && Asymp.\ Std.\ error = 0.004879 \\
&\hat{\sigma}^2_{w\beta_1} = 0.00188 && Asymp.\ Std.\ error = 0.001343 \\
&\hat{\sigma}^2_{w\beta_2} = 0.0123 && Asymp.\ Std.\ error = 0.00844
\end{aligned}$$

Finally we plot the smoothing estimates of the beta coefficients for the two regressors: BTP interest rate and CPI rate.

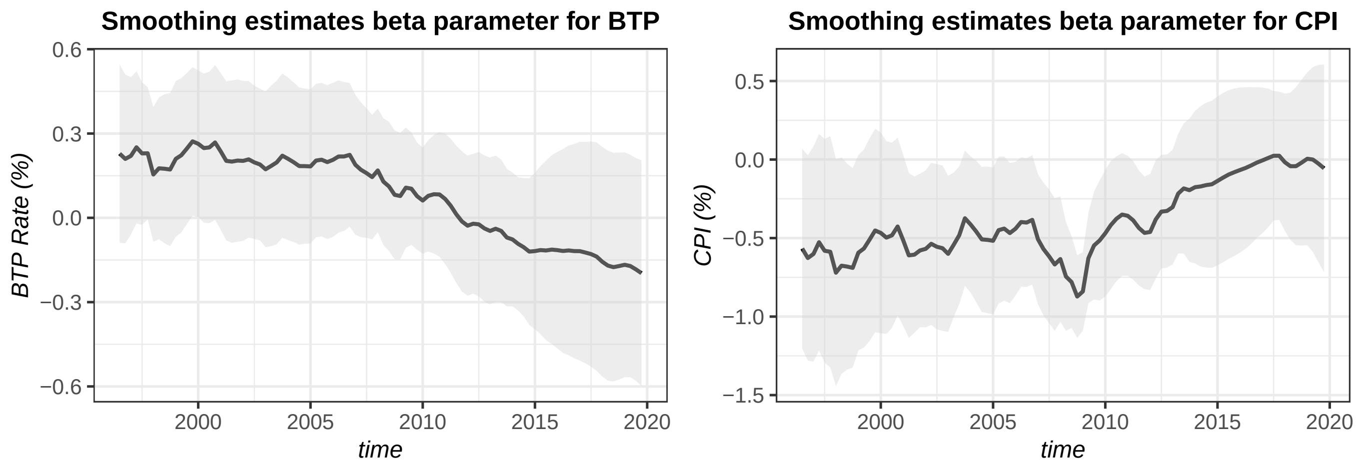


Figure 4. Plot of smoothing estimates of beta coefficients

These two plots are, in our opinion, really interesting and informative since they are describing how the relationship between GDP growth rate and BTP interest rate and between the GDP growth rate and CPI rate has changed over time. We can notice, as we expected, that the size of these coefficients had decreased from the financial crisis on, but unexpectedly for the coefficient of BTP we also register a change in sign from 2010, i.e. there is now a negative relationship between BTP and GDP. In particular we can see that there is an increase in the magnitude of this coefficient, which has reached a value of -0.2 in 2019, and does not seem to be decreasing any time soon. On the other hand the β for the CPI rate goes from -0.75 to near zero, showing an abrupt change during the financial crisis. We have that inflation has now an almost insignificant relationship with the GDP growth rate.

Performance Indicators

As a final step we show the most common performance parameters to compare how the models perform in term of forecasting accuracy. The Mean Absolute Error (MAE) and the Mean Squared Error (MSE) express the average model prediction error in terms of units of the variables used (in this case, the GDP change in percentual), while the Mean Absolute Percentage Error (MAPE) expresses the model prediction error in absolute percentages.

Errors criteria	Random Walk + Noise (2)	Linear Regression Model (3)	Linear Regression Model (4)	Bayesian linear regression (5)	Dynamic linear regression (6)
MAE	0.4264929	0.4329216	0.4376340	0.5533037	0.4824994
MSE	0.3682446	0.3418759	0.3470332	0.8953326	0.7816666
MAPE	6.3193047	2.7739905	2.8171182	4.9807622	6.4564172

From these results we can conclude that the best predicting model is the linear regression model. In particular, the model with three regressors (3) is slightly better than the model with two regressors (4). However, we have to keep in mind that these performance indicators do not penalize for the number of regressors. For this reason, during the evaluation stage, we have used the Information Criteria, which suggests a preference for the linear regression model with two regressors (4). Moreover, considering that the three error indicators can also be in contrast among themselves as we can see for the Random Walk plus Noise and Linear Regression Model. In fact, MAE seems to suggest that the Random Walk plus Noise (2) model is the best model, while MSE and MAPE seems to suggest that the Linear Regression Model is better (3). A last remark is that the fact that the best model is the Linear Regression Model may suggest a possible stationarity of the GDP growth rate.

Final Conclusion

To sum up, the results produced in this project are interesting even if further developments could be introduced into the analysis. We started with an HMM to describe the discrete latent process underlying the GDP growth rate, finding the results consistent with the ones of the unemployment rate. This indicates that the model implemented captures the abrupt change points that occur in the Italian economy. However, we developed a more complete model considering the

DLM class of state space models. After the implementation of a simple random walk plus noise model, we construct a static and then a dynamic regression trying to explain how monetary variables impact the GDP growth rate. In this final part of the project we find particularly interesting the variation of the beta coefficients over time, since they could give an idea on how these nominal variables were related to the GDP growth rate before and after the crisis. In conclusion, our results are not always consistent, indeed some diagnostic checks for the residuals and the discording results between the information criteria and the various kind of performance indicators are sometimes not encouraging. For example, as we have seen the random walk plus noise model seems to have better performances and this could be a symptom that the relation we are trying to analyze could not be captured by a linear regression model. There are different reasons for this lack of perfect description of the economic variables and maybe we should check better the assumption underlying the entire analysis like the non stationarity of the time series, to maybe try to transform it in a stationary one and implement an ARMA model or other model which will lead to more efficient results. However, the analysis is a starting point to further develop the adoption of these models to study economic variables.

References

Campagnoli P., Petris G., Petrone S. (2009). *Dynamic Linear Models with R. New York*: Springer-Verlag

Chatfield C., Xing H. (2019). *The Analysis of Time Series. An introduction with R*: CRC Press

Petris G. *An R Package for Dynamic Linear Models*: October 2010, Volume 36, Issue 12. Journal of Statistical Software

Visser I., Speekenbrink M. *depmixS4: An R Package for Hidden Markov Models*: August 2010, Volume 36, Issue 7. Journal of Statistical Software

Zucchini W., MacDonald I., Langrock R. (2016). *Hidden Markov Models for Time Series. An introduction using R*: CRC Press