Portuguese Consumer Price Index

Catarina Pinto nº48397, João Santos nº48323, Leonor Santos nº51983, and Mariana Marques nº48340

[1] ISEG; Rua do Quelhas 6, 1200-781 Lisboa
[2] info@isegexecutive.education
https://www.iseg.ulisboa.pt/pt/

Abstract. In this project our main goal is to analyse which variables have a greater impact on inflation, for that, we analysed the inflation relative to the period between 1960 to 2022, with monthly data, as well as made a prediction model that would fit the variables. In order to do so we resorted to Jupyter Notebook and we resorted to INE for collecting data.

Keywords: Inflation · Time Series · Prevision Models.

1 Method

1.1 Model Construction

The definition of inflation, according to European Central Bank, "Inflation occurs when there is a broad increase in the prices of goods and services, not just of individual items; it means, you can buy less for €1 today than you could yesterday. In other words, inflation reduces the value of the currency over time.".

In order to meet our goal we are going to use a linear regression model to compute the correlation between the variables presented in the following table:

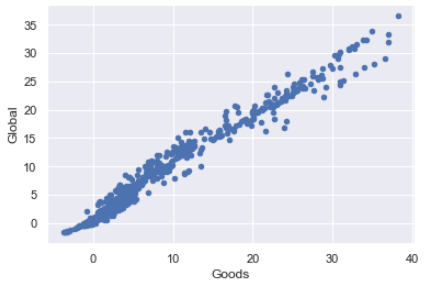
Variables	Description	Unit
Global	Inflation Value/CPI	%
Goods	Retail Products	%
Services	A system supplying a public need	%
Unprocessed foods	Foods that contain no artificial flavors or synthetic ingredients	%
Energetic Prices	Prices of the Energy	%

Table 1: Variables.

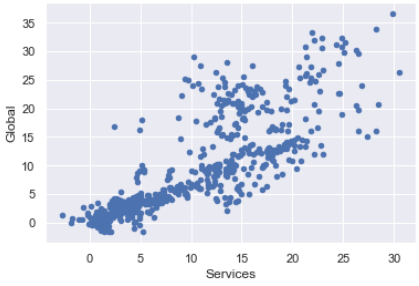
In this context our dependent variable will be the "Global" and the other variables represented are the independent ones.

2 Results

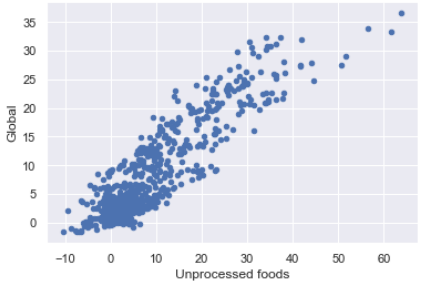
2.1 Correlation Analysis



(a) Correlation between Goods and Global



(b) Correlation between Services and Global



(c) Correlation between Unprocessed Foods and Global

(d) Correlation between Energy Prices and Global

Fig. 1: Correlation between the Independent Variables and the Dependent

Now let us analyze the correlation between our independent variables and inflation, more precisely for the variable goods and the variable services, which we observe as the extremes of our model.

In the first graph, we can observe an almost perfect positive correlation between the 'Goods' and the 'Global', which means that the 'Goods' have a strong impact on inflation. To sustain this statement, we did the correlation between all independent variables with the dependent one and observed a correlation of 0,986937 (98,69%). Moreover, we also did the VIF and obtained the following 21.177 complementing what we previously stated.

In the last graph, we notice there is a less strong positive correlation between the variables (energetic prices and inflation) when compared with the first case analyzed. In this context, the independent variable has a lesser impact on the dependent one. This result is, once again, sustained by the obtained correlation value of 0,8085 (80,85%), as well as a VIF of 5.410.

Complementary to this analysis, we have some comments on jupyter notebook about the VIF and the corresponding classifications, as well as some observations on the correlation graphics.

2.2 Time series

Fig. 2: Time series from January 1960 to October 2022 of the Inflation.

We can observe that until August 1964 there wasn't heavy inflation, in the sense that we can see a small to medium variation, so naturally, there weren't any extreme values during this period.

However, afterwards we observed that there is a highly volatile inflation period, reaching one of our maximum values, 36,66% in May of 1977. This was caused by an oil crisis that started in 1973, to aggravate the situation, in April 1974, Portugal went through a revolution, creating very unstable governmental conditions. So, from December 1972 until the mid-1980s we see an instability in the values as well as a fast-growing of them. In line 300, corresponding to November 1984, we can observe a decrease in these values, which may indicate possible stability.

Although the values observed were still high (since according to BCE the limit of inflation should be 2% or below), we could observe a "constant" decrease in the values until they stabilized in the periods afterwards. One of the causes for this may be that in 1986 Portugal joined the European Union, and in January 2002 Portugal's currency became the Euro. All of this caused inflation to decrease since there are specific requirements to join these organizations.

2.3 Comparison with each independent variable

In this section, our focus is to see if the volatility of the prices in each independent variable has the same behaviour as the volatility of the dependent variable.

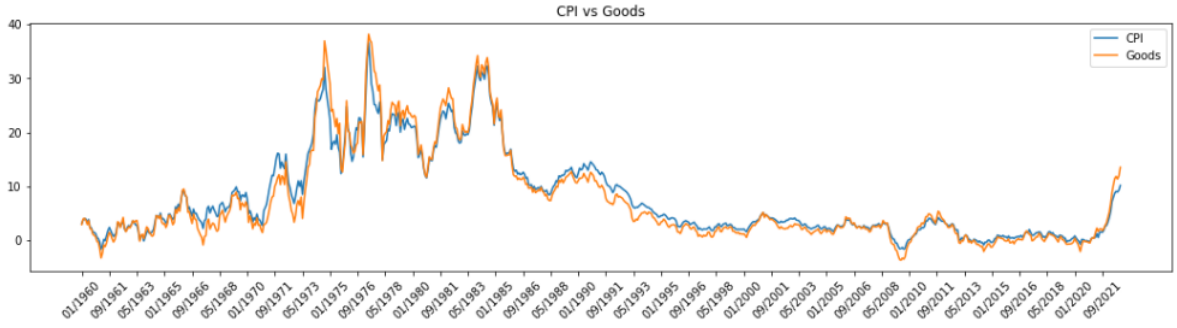


Fig. 3: Comparison with the Inflation and the Goods.

In this case, we can observe that the variable "Goods" has much similar behaviour to some of the following ones. A possible explanation is that the market price of general goods highly influences inflation.

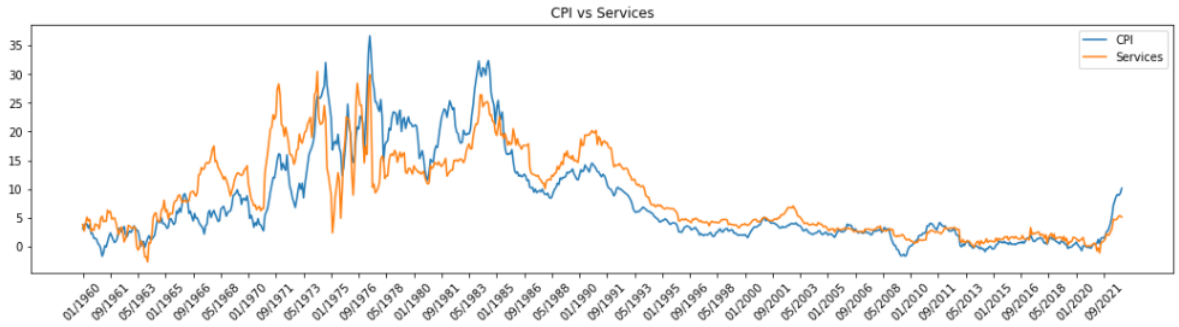


Fig. 4: Comparison with the Inflation and the Services.

However, the variable "Services" doesn't seem to have as much of a similar behaviour to inflation, however, in 1993 we observe the behaviour of the services be more similar to the inflation. Although the services may affect the prices of the services, these tend to adjust their values to the demand that they have. For example, we can observe that when the inflation tends to be higher the services tend to be lower, since with the inflation people usually lose purchasing power so they don't have the ability to buy a specific service, however, in the long run, we can observe that the behaviour of the services tends to match with inflation.

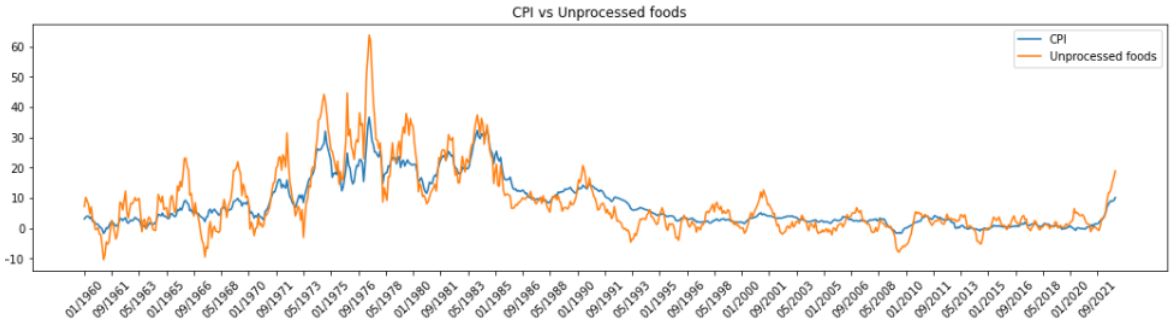


Fig. 5: Comparison with the Inflation and the Unprocessed foods.

In the case of the variable "Unprocessed foods", don't follow a very rigorous behaviour when compared to the inflation. Since unprocessed foods are more influenced by external conditions, for example, metrological conditions, it's natural that their behaviour isn't equivalent. To better understand, the Harvard School of Public Health, defined "Unprocessed foods include the natural edible food parts of plants and animals", so the only way that inflation could affect is the feedstock that normally farmers need for their plantations or to create animals. Nevertheless, unprocessed foods still have a strong correlation with inflation, which means that their behaviour will be very close, although not perfect, to inflation.

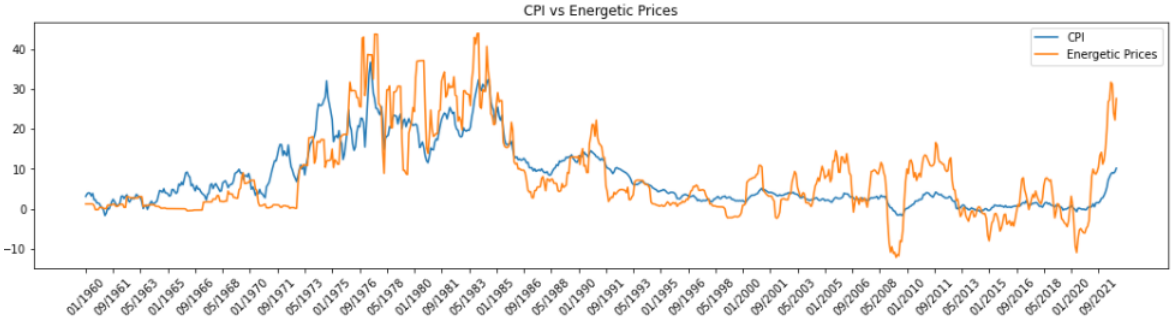


Fig. 6: Comparison with the Inflation and the Energetic Prices.

The "Energetic Prices" have less similar behaviour with the inflation, since this variable is very influenced by macroeconomic factors, for example, if we focus on the period between 2021 and today, we see an increase of the values which may be correlated with the war that we are experiencing in 2022 in Europe and consequences of the pandemic crisis. Another factor that may explain this kind of independency is that the energetic prices could have some effects on inflation

since if these increase then the prices of the other's products tend to have the same behaviour because most of them use energetic products as intermediate products.

So, after this analysis we can observe that inflation has a very strong correlation with all the independent variables, as we have seen before since the independent variables can influence them directly and indirectly.

2.4 Predictions Models

As stated before, we intend to make an accurate prediction model. As such we have computed some prediction models that we analyzed.

In these models we used the independent variable of "Energetic Prices" to predict our dependent one ("Global"). In our model, we're allowed to do this because the independent variable that we've chosen has a strong correlation with our dependent one

OLS Model For the first prediction, we use a linear regression model. It is critical to evaluate whether this is a good prediction model for our data. For that, we do a goodness-of-fit measure, like computing R-squared. In this case, we can observe that, by analyzing the R-squared score, this is in fact, a good prediction model, although not the best one seen as our values for R-squared is a little lower than we would prefer (R2 score (train) = 0.65 and R2 score (test) = 0.68) meaning a less precise prediction).

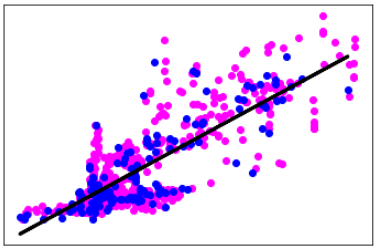


Fig. 7: OLS Graphic Representation for the Energetic Prices and Global. Pink dots-train. Blue dots-test

MLP Model For the second prediction, we used a multilayer perceptron model. Computing once again the value of R-squared, we see that very similar to the previous model, the MLP model is a good prediction model, but not the best one. (R2 score (train) = 0.65 and R2 score (test) = 0.68) meaning a less precise prediction).

Random Forest Model For our final prediction, we used the Random Forest model. In this model, we obtained a better model in terms of goodness-of-fit measure for the R2 score (train) (0.92), however for the R2 score (test) we obtain a less accurate prediction (0.58), more similar to the previous models.

In sum, for the R2 score (train) we see that the best prediction model would be the Random Forest model, however for our goal, that is the prediction, we should focus on the other's two models since they have a better R2 score (test).

3 Conclusion

As we expected all the independent variables have a high correlation with the dependable one since they have mutual effects on each other. In terms of prediction, we could observe that the OLS model and the MPL model are better since they represent higher R2, however for the R2 score (train) the Random Forest model is the better one.

References

1. Vieira, F. E. P.: O impacto do preço do petróleo na economia e na bolsa de valores portuguesa. , 11–24 (2020)
2. Harvard T.H. Chan. School of public Health,https://www.hsph.harvard.edu/nutritionsource/processed-foods/. Last accessed 22 Nov 2022
3. 2021portugal.eu,https://www.2021portugal.eu/pt/noticias/o-percurso-de-portugal-na-uniao-europeia/. Last accessed 22 Nov 2022
4. INE Homepage,https://www.ine.pt/xportal/xmain?xpgid=ine_mainamp;xpid=INE. Last accessed 12 Nov 2022