Predictive Analytics

Analyzing and Forecasting Inflation Rates in Hungary, Italy and Spain

COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Jenő Tóth (158386)

Character count: 16,545

Page count: 10

Submission date: August 18, 2023

Predictive Analytics Written Exam

MSc. in Business Administration and Data Science

Introduction

Inflation is the rate of increase in prices over a given period of time, diminishing the purchasing power of money (Oner, 2022). Slight positive inflation rate is desirable for monetary institutions, as there may be unfavorable outcomes caused by deflation, such as in the case of Japan. Despite this, inflation has negative connotations, and when it gets high, it causes problems not only for central banks, but for everyday people as well. There may be a number of different reasons why inflation can drastically increase, these usually come in the form of an economic shock, such as the pandemic or the energy crisis catalyzed by the war in Ukraine.

Due to recent shocks, inflation has become one of the most important topics in a number of countries around the world. While policy interventions may be applied to help lower inflation, oftentimes it converges to a smaller value due to its nature. Monthly inflation values are calculated based on the same month a year before, meaning that if there was already high inflation in the past, the present values may be lower, due to the absorption of the previous shocks. If past inflation values are related to present and future ones, it is possible to build forecasting models to predict how inflation can change in the future. This paper focuses on building such models. The paper analyzes yearly inflation data between 1970 and 2022 for Hungary, Italy, and Spain (Ha, 2021), and uses Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR) models for forecasting the countries' inflations up until 2030.

The three countries were picked due to their similarities, but also their key differences. Currently, all three countries are European Union (EU) member states, however, they joined at different times. This allows us to examine how their inflation rates change around the join dates. Italy and Spain also use the Euro as their currency, Italy introducing it in 1999, while Spain adapting it in 2002. Hungary, on the other hand, uses Forint, which may lead to more unstable inflation compared to the Euro area (Álvarez et al., 2019).

Methodology

The research used two forecasting models: ARIMA and VAR. ARIMA is a popular time series forecasting methodology. The technique combines both autoregression (AR) and moving average (MA) models, taking into account the differences to make the time series stationary. The AR component looks at the relationship between an observation and its predecessor, while the MA component looks at an observation and the residual error from the previous forecast. The model therefore looks at 3 parameters, denoted by p, d, and q. “p” refers to the number of lags to include, “d” determines how many times the data has to be differentiated in order for it to be stationary, and “q” determines the number of lagged forecast errors to be included.

While ARIMA is used for a single time series, VAR models the interdependencies and dynamics across multiple time series. This can help us understand the relationships between the inflation rates in the three analyzed countries. Each variable in a VAR model is modeled as a linear combination of past lags of itself and past lags of the other variables, making it apt for analyzing the dynamic impact of changes in one time series on another over subsequent periods.

Before creating the models, we have to take a number of steps. After loading the necessary R packages and the dataset, the three observed countries should be selected. The data is also horizontal, with each year being a separate column. This is transformed into longitudinal format, with “year” being a single column for the time variable. We now have four main columns: year, country, inflation type, and inflation value. There are 5 different types of inflation to explore. Headline Consumer Price Inflation (HCPI) measures the total inflation in a country, Energy Consumer Price Inflation (ECPI) looks the changes in energy prices, Food Consumer Price Inflation (FCPI) measures the inflation rate of food products, Official Core Consumer Price Inflation (OCCPI) measures inflation without including ECPI and FCPI, and Producer Price Inflation (PPI) measures the rate at which prices of goods change at the wholesale level. For HCPI, ECPI and FCPI, there are no missing values. For OCCPI, data is only available from 1976 for Spain and 1991 for Hungary. For PPI, there are missing values until 1972 for Spain, 1979 for Hungary, and 1975 for Italy. These missing values were not included in figures, and for the forecasts, they were replaced by zeros. The following figure shows the values for the different types of inflations in the research.

Figure 1: Rates of different inflation types for the three observed countries

After preprocessing the data, two functions were built, one for the ARIMA models, and one for the VAR models. For the ARIMA analysis, 15 different scenarios were created, based on the combinations of the 5 inflation types and the 3 countries. The function therefore had three arguments: country, inflation type, and start year, which was manually set to be the first year from which data was available in the given category.

After filtering for the given country and inflation type, the function plots the Autocorrelation Function and the Partial Autocorrelation function in order to gain a better understanding about the data, and whether it needs differentiating. Next, the function runs a series of Augmented Dickey-Fuller (ADF) tests to see how many times we need to differentiate the data to make it stationary. While we cannot reject the ADF test's null hypothesis (meaning that the time series is non-stationary), we differentiate it an additional time, until the series is stationary. A "d" parameter is then stored, its value being equal to the number of differentiations needed.

The ARIMA model is created after this, finding the optimal p and q parameters to join our already determined d parameter. This is done by calculating the Akaike Information Criterion (AIC) for each possible parameter combination, and picking the model where the AIC is the smallest. This gives us the model that we can use for forecasting a given set of data. This forecast is then plotted and finally the model is evaluated. For the evaluation, we split the data we have into 80% training and 20% testing. A new forecast is created for the training set, and it is validated by using the test set. We can then measure a number of errors, such as the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE) or the Mean Absolute Percentage Error (MAPE).

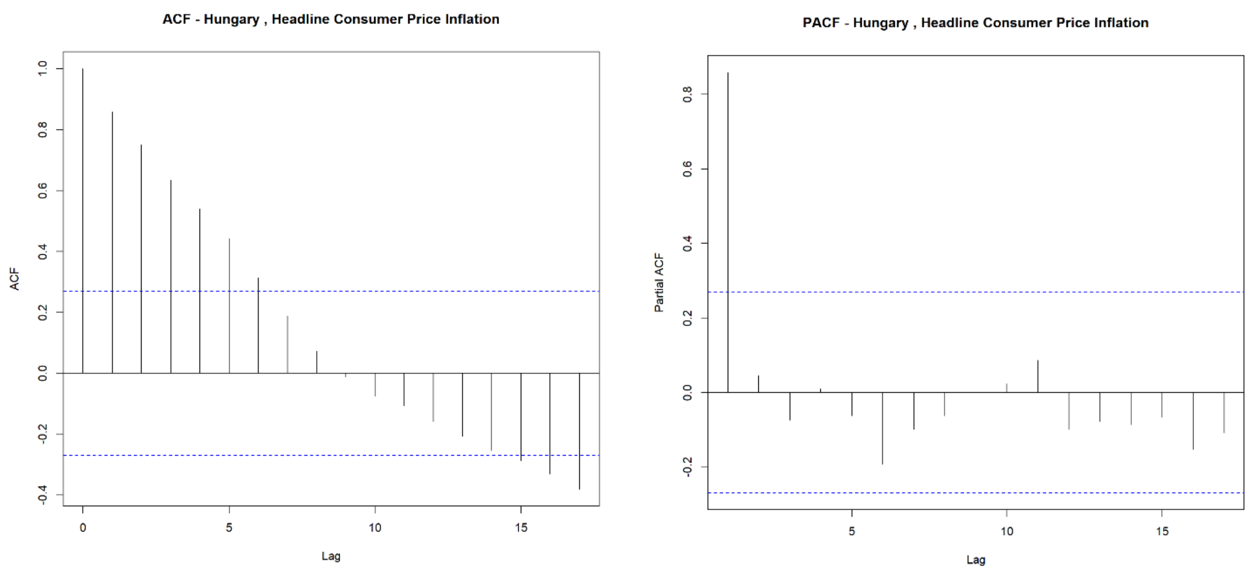


Figure 2: An example of the Autocorrelation Function and the Partial Autocorrelation Function in the ARIMA model

The VAR function works similarly to the ARIMA function. For the Vector Autoregression (VAR) analysis, the method examines how one variable (like inflation in Hungary) might impact another (like inflation in Italy). A vital difference between ARIMA and VAR is that while ARIMA is for univariate time series data, VAR can handle multivariate series.

The first step of the VAR function is to filter out the data for the specified inflation type from the long-format dataset, and then pivot the data into a wider format with one column for each country. This results in a time series dataset where each row represents a year, and each column is the inflation for a particular country. Given that the VAR model often requires the series to be stationary, we difference the data for each country. Differencing is a common method to make a time series stationary, which involves subtracting the current value from the previous.

With our data preprocessed, we proceed to model selection. Two VAR models are estimated based on two different criteria: the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria help determine the optimal number of lags to include in the model. Both models' details are reported to offer insights into their coefficients, residuals, and overall fit. Residual diagnostics are essential. We analyze the residuals of the AIC-selected model using the Ljung-Box test, which tests for autocorrelation in the residuals. Ideally, a well-fitting model should have no significant autocorrelation in its residuals.

Granger causality tests follow, determining the directional relationships between the countries. In essence, a Granger causality test helps ascertain if past values of one variable (for example Italy's inflation) provide information about the future values of another variable (for example Hungary's inflation). Impulse response functions (IRFs) are then plotted. IRFs trace the effect of a one-time shock in one of the variables on the other variables in the system. For instance, the plots may depict the effect of a sudden increase in Hungary's inflation on Italy's inflation over time. A Forecast Error Variance Decomposition (FEVD) plot is generated, providing insight into the proportion of the movements in a sequence due to its own shocks versus shocks to the other variable.

Lastly, similar to the ARIMA analysis, we evaluate the accuracy of the VAR model by splitting the data into training and testing sets (80% and 20%, respectively). After fitting the VAR model to the training data, we then test its predictive accuracy on the test data, which can be assessed with measures like RMSE, MAE, and others. Overall, the VAR function encapsulates a comprehensive analysis of the relationships between inflation rates in three countries, providing insights into causality, impacts of shocks, and forecasting prowess.

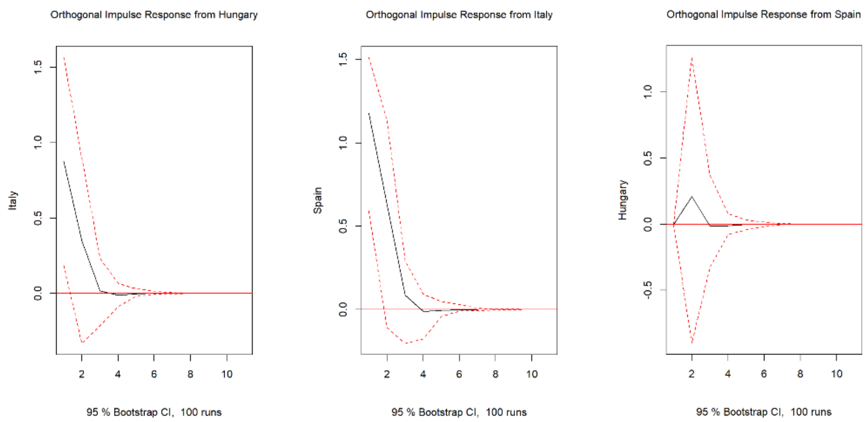


Figure 3: Examples of Orthogonal Impulse Plots between the three countries

Results

The following table summarizes the 15 different ARIMA models and their results.

Model Name	Parameters	AIC	BIC	RMSE	MAPE
Hungary - HCPI	ARIMA(2,0,0)	306.89	314.77	4.38	675.18
Italy – HCPI	ARIMA(1,1,1)	235.93	241.79	2.57	1275.3
Spain – HCPI	ARIMA(2,1,0)	232.37	238.22	2.86	591.57
Hungary - ECPI	ARIMA(1,0,0)	421.26	427.17	11.33	746.61
Italy – ECPI	ARIMA(1,0,0)	401.9	407.81	14.76	1697.55
Spain – ECPI	ARIMA(1,0,1)	365.43	373.31	11.16	301.09
Hungary - FCPI	ARIMA(2,0,3)	348.46	360.28	7.6	510.48
Italy – FCPI	ARIMA(1,1,0)	244.4	248.3	2.72	537.48
Spain – FCPI	ARIMA(1,1,0)	268.55	272.45	2.92	103.78
Hungary - OOCPI	ARIMA(1,2,3)	165.51	172.52	9.95	212.16
Italy – OOCPI	ARIMA(0,1,1)	409.22	413.12	10.11	346.17
Spain – OOCPI	ARIMA(1,1,1)	370.41	375.9	9.1	3024.55
Hungary - PPI	ARIMA(1,0,0)	292.85	298.21	6.93	326.42
Italy – PPI	ARIMA(1,0,5)	415.46	430.42	17.08	415.79
Spain – PPI	ARIMA(2,2,4)	325.33	338.43	11.4	822.94

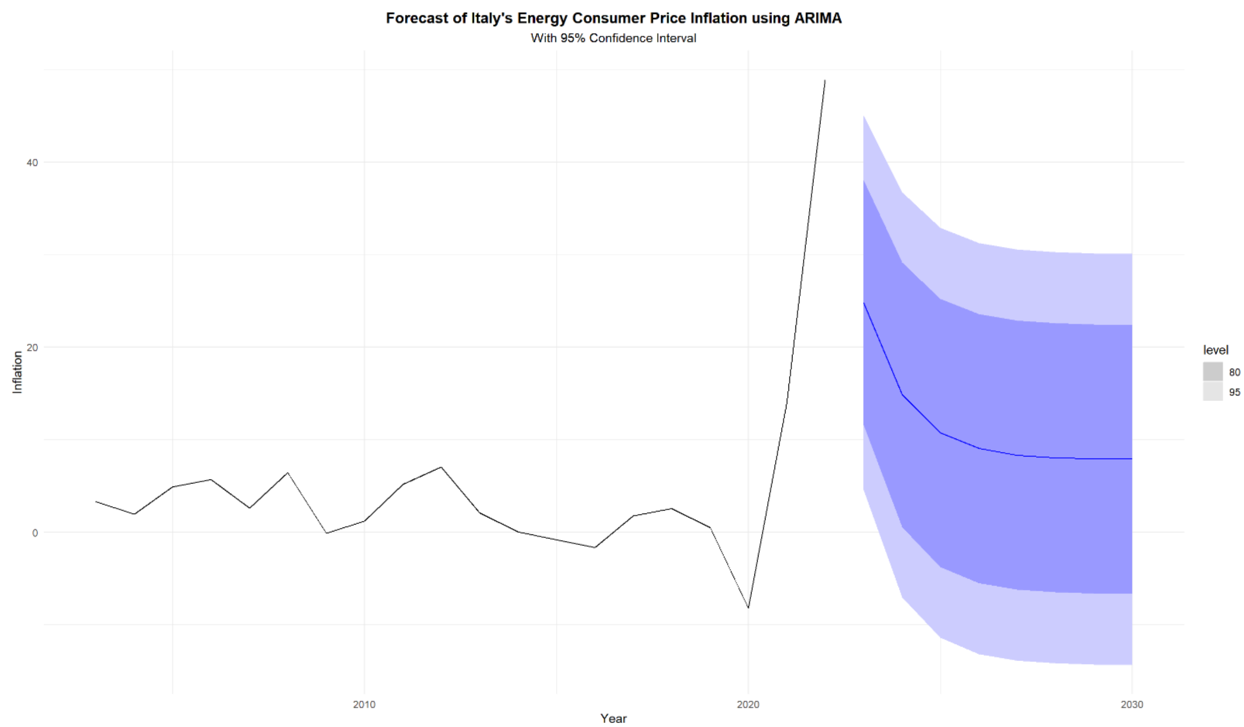


Figure 4: Italy's Energy Consumer Price Inflation forecast using ARIMA.

The other forecasts may be found in the Appendix

For the VAR models, 5 models were created to see the countires' interdependencies in each inflation type. The following table summarizes the models.

Model Name	Parameters	AIC	BIC	RMSE	MAPE	Granger Causal?
HCPI	VAR(5)	680.68	780.59	4.37	129.2	No
ECPI	VAR(5)	1029.39	1129.3	14.78	143.1	No
FCPI	VAR(5)	775.75	875.66	2.26	173.72	No
OCCPI	VAR(5)	1119.26	376.76	12.86	920.81	Yes
PPI	VAR(5)	1019.16	1119.07	21.78	2524.6	Yes

Discussion

As we can see, forecasting inflation with ARIMA and VAR provides us with mixed results. The confidence intervals for the ARIMA model are too large to provide any fruitful conclusions. This may be caused by the fact that the data is annual, not more common, such as monthly inflation. Still, we can see that our models were better at predicting some types of inflation than others.

When taking a deeper look at the ARIMA models, the first takeaway can be that despite the data being non-stationary in most cases, oftentimes the best results were achieved when the “d” parameter was set to 0.

For the ARIMA models, Hungary's forecasts generally showed the least precision when compared to Italy and Spain across most inflation types. Taking HCPI as an example, the RMSE for Hungary's HCPI forecast is 4.38, which is notably higher than Italy's RMSE of 2.57 and Spain's RMSE of 2.86. This trend suggests that the inflation data for Hungary might be more volatile or perhaps the model parameters for Hungary may not be capturing the underlying data dynamics as effectively as they do for Italy and Spain.

Similarly, in the ECPI category, Hungary's RMSE is 11.33, while Italy and Spain register values of 14.76 and 11.16 respectively. This indicates that while Hungary's forecast for ECPI is closer in accuracy to Spain's, Italy's ECPI forecast seems to be the least accurate. It's also worth noting that MAPE values, another measure of forecast accuracy, show similar trends with Hungary typically having higher values, indicating a larger percentage error in predictions.

VAR models were used to see if there are interdependencies between countries in a given inflation type. The idea behind interdependencies is to understand if past values of one variable can help us predict current or future values of another variable. In our case, the question is whether we can use one country's inflation data to predict another one's. One of the standard methods to assess such temporal relationships is the Granger causality test. Granger causality tests look at two models: a baseline model that only uses one country’s inflation for prediction, and another one which combines it with the inflation values of another country. If the second model performs better than the first, we can say that there is Granger causality between the two variables.

This paper uses a p-value of 0.05 as the threshold for Granger causality. With this cutoff, we can see significant results in two cases. The first is predicting Italy's OOCPI with the help of Spain's data (p-value 0.039) and the second is predicting Italy's PPI together with Spain's PPI values (p-value 0.012). One explanation why Hungary does not appear is that it uses a different currency, meaning that economic indicators may not move as much in sync as for two eurozone countries.

Regarding the accuracy of the VAR models, their AIC values can provide some insight. A lower AIC value indicates a better model fit, taking into account both the goodness of fit and the number of parameters. For instance, the HCPI VAR model has an AIC of 680.68, the ECPI model has 1029.39, and so on. While it's not strictly accurate to compare AIC values across different datasets, within the context of these VAR models, it's clear that certain inflation types (like HCPI) have a better model fit than others.

Furthermore, the RMSE values from the VAR models shed light on the models' predictive accuracy. The RMSE is a measure of how far off the model's predictions are from the actual values, with lower values indicating better accuracy. In this case, the HCPI VAR model has an RMSE of 4.37, ECPI's is 14.78, and so forth.

Conclusion

Forecasting inflation remains a difficult task, as it may be influenced by a number of factors both within and beyond a country's borders. This paper examined inflation trends across Hungary, Italy, and Spain using ARIMA and VAR model. It examined Headline Consumer Price Inflation, Energy Consumer Price Inflation, Food Consumer Price Inflation, Official Core Consumer Price Inflation and Producer Price Inflation between 1970 and 2022, forecasting until 2030. Due to the use of annual data, the results can give us valuable insights, but they are inconclusive.

The ARIMA models showcased a somewhat counterintuitive tendency: despite the non-stationary nature of much of the data, models with no differencing often provided optimal results. This underscores the need for careful model selection and validation when working with time series data. Disparities in forecasting precision across the three nations, especially Hungary's general lower accuracy, further highlight the unique economic dynamics each country exhibits. The different economic histories of the three countries, such as Hungary’s socialist regime and the countries joining the EU at different times may be answers to these disparities.

To see whether one country’s inflation data can help us understand another country’s values, VAR models were built, testing whether or not there is Granger causality in our data. The paper found little evidence of this, and only among the two countries who have similar economies and use the Euro.

For future research, augmenting these models with external variables describing the economies in more detail might provide deeper insights. The inclusion of other countries can also help us find patterns about which countries can be useful for predicting a given economy’s inflation. Additionally, the use with higher frequency data, when available, could further improve the predictions.

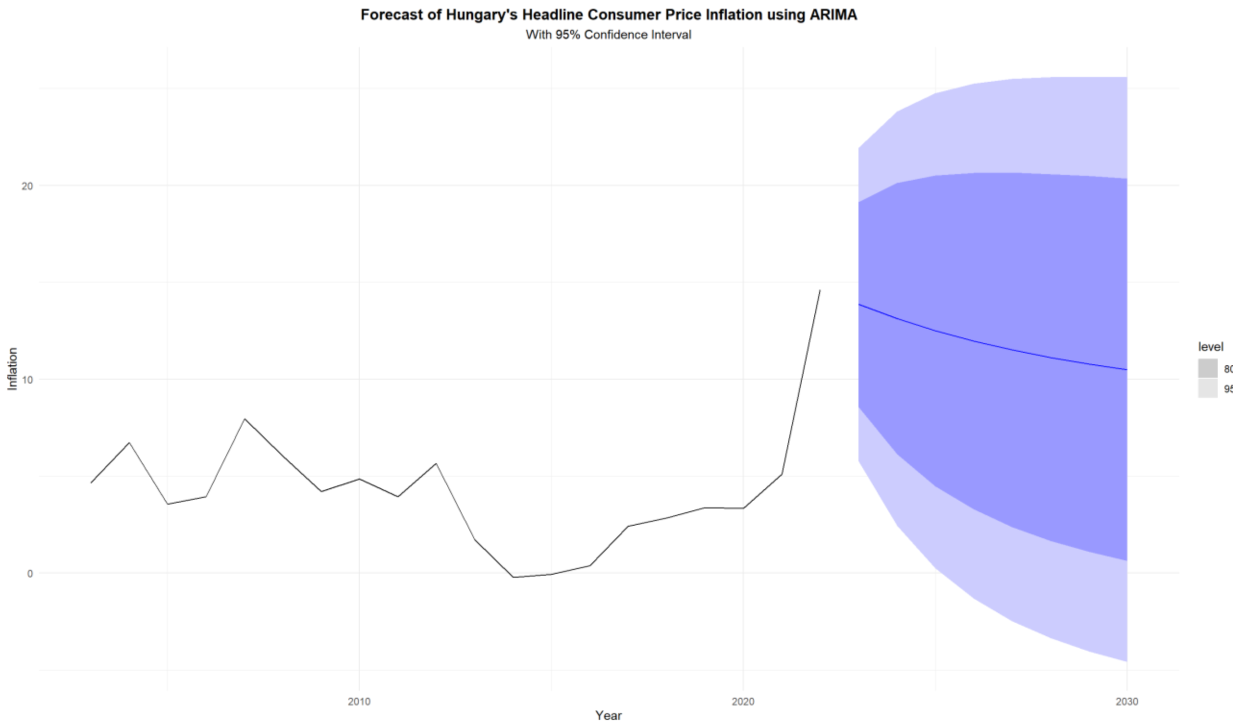
References

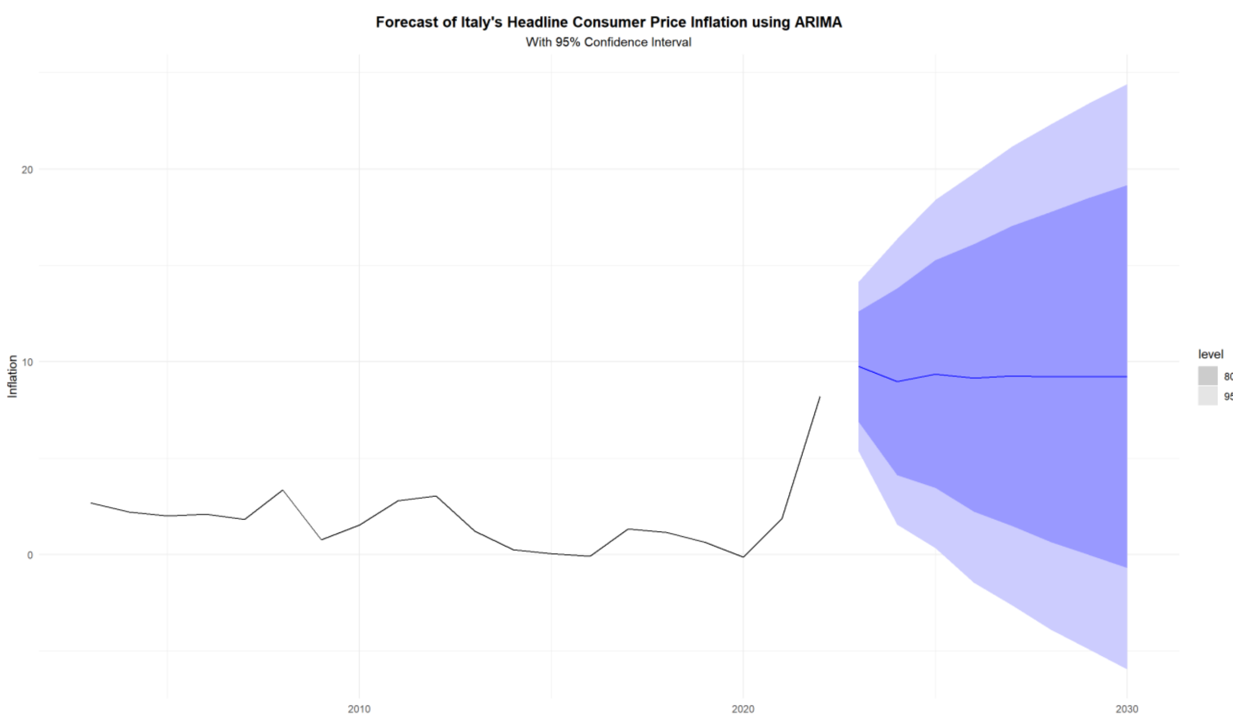
Álvarez, L. J., Gadea, M. D., Gómez-Loscos, A. (2019). Inflation interdependencies in advanced economics. Banco de España.

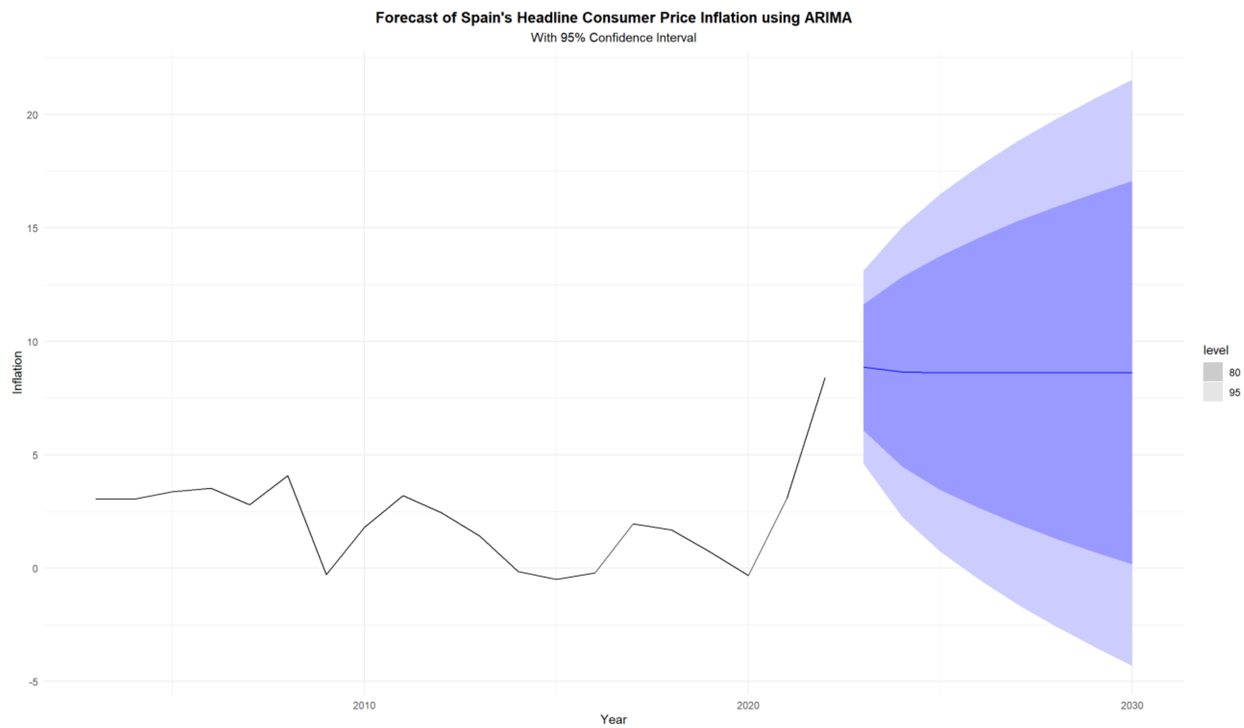
Ha, J., Kose, M. A., Ohnsorge, F. (2021). "One-Stop Source: A Global Database of Inflation." Policy Research Working Paper; No. 9737. World Bank, Washington, DC."

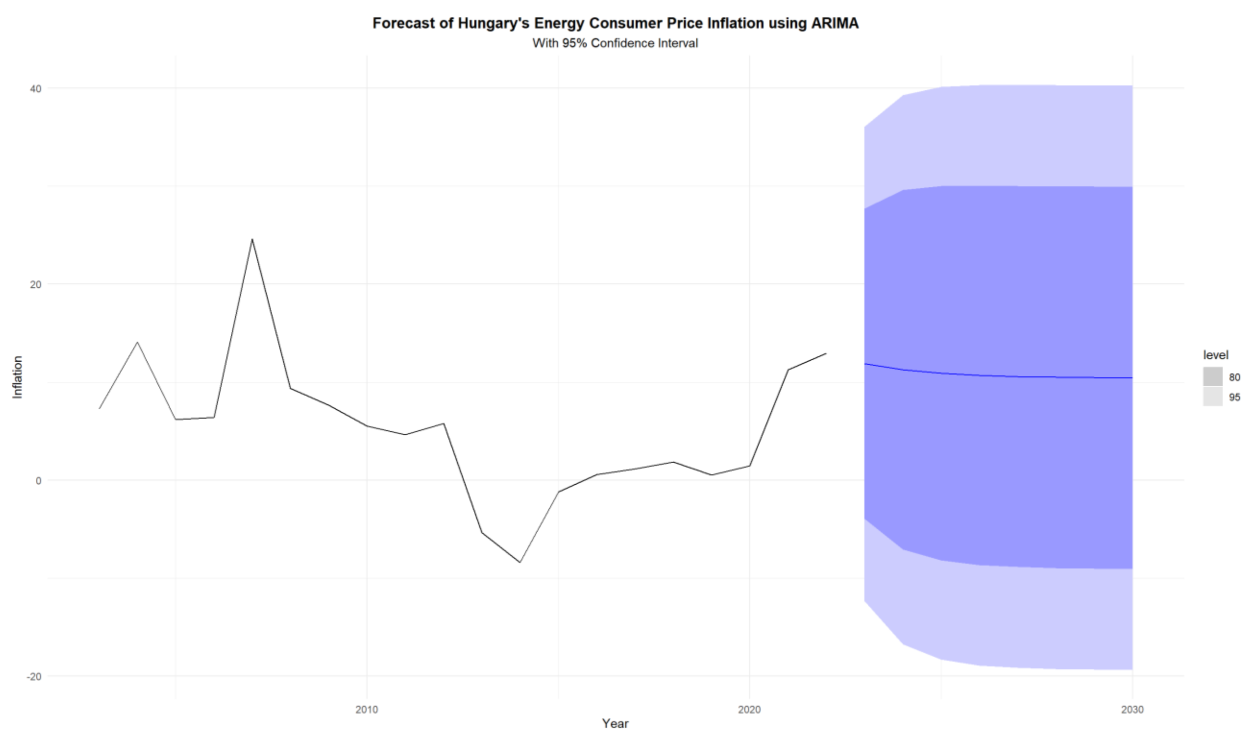
Oner, C. (2022). Inflation: Prices on the Rise. International Monetary Fund. https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/Inflation

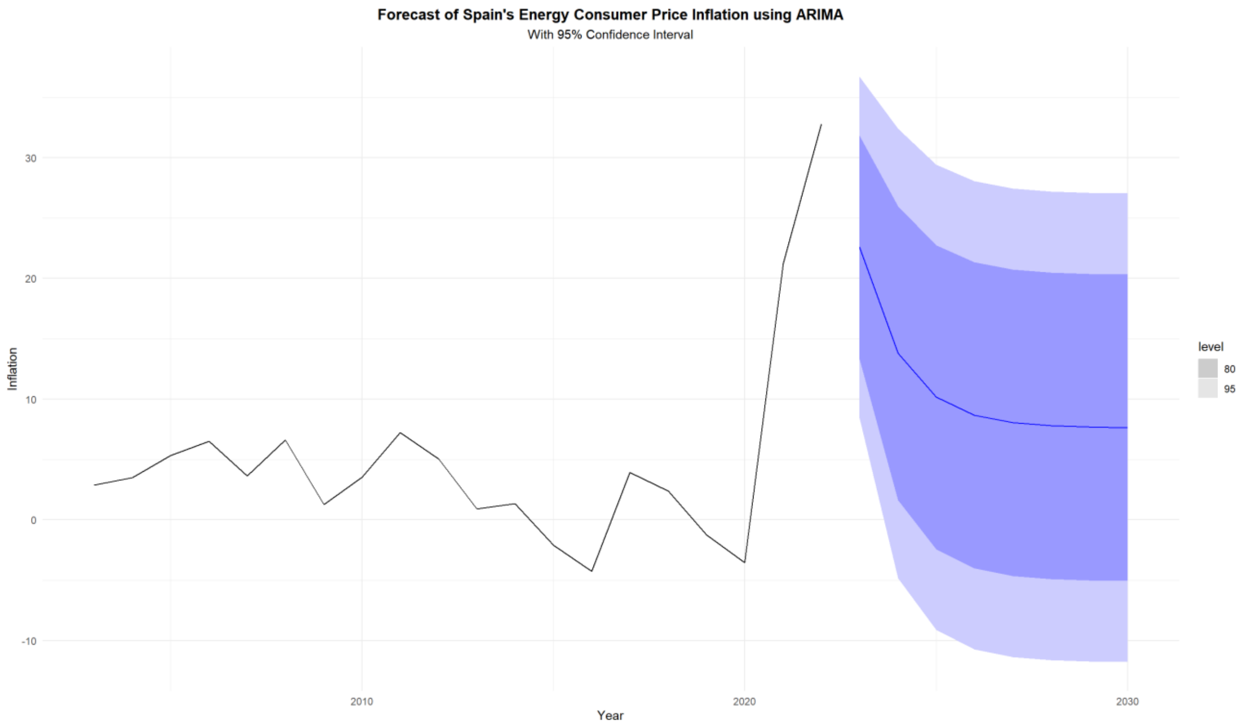
Appendix

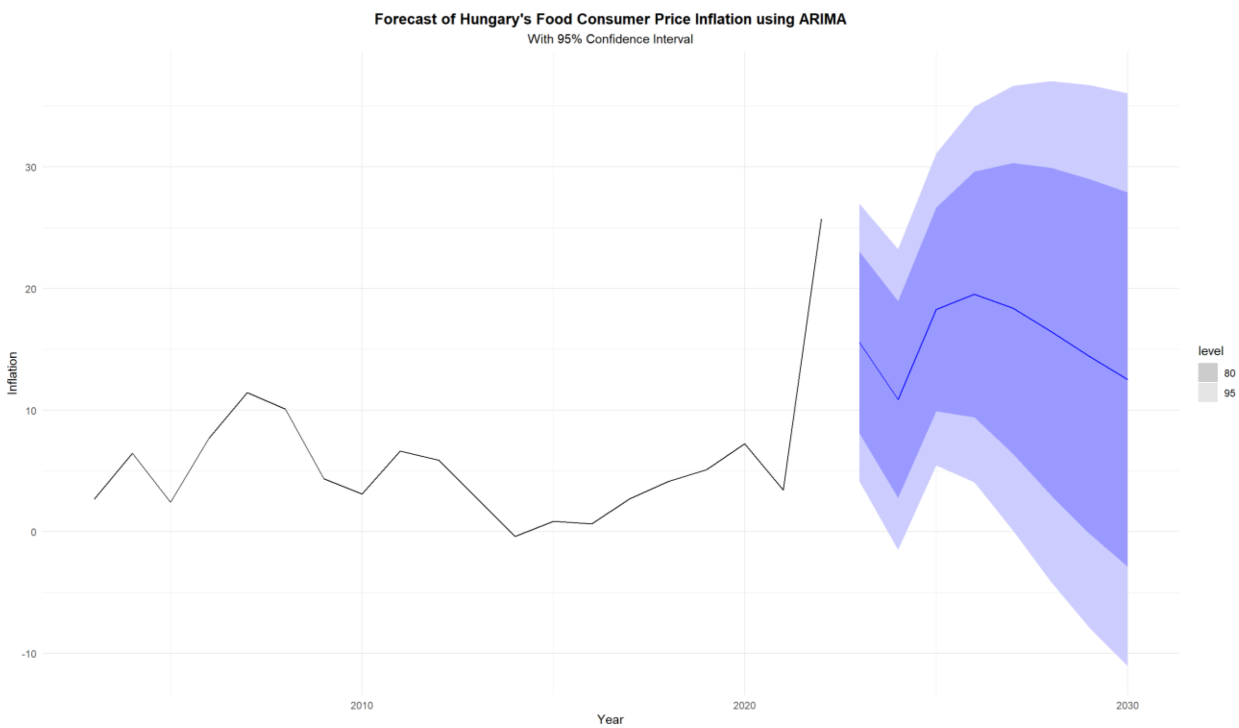

Forecast of Hungary's Headline Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
20
10
0
2010
2020
2030
Year
level
80
95

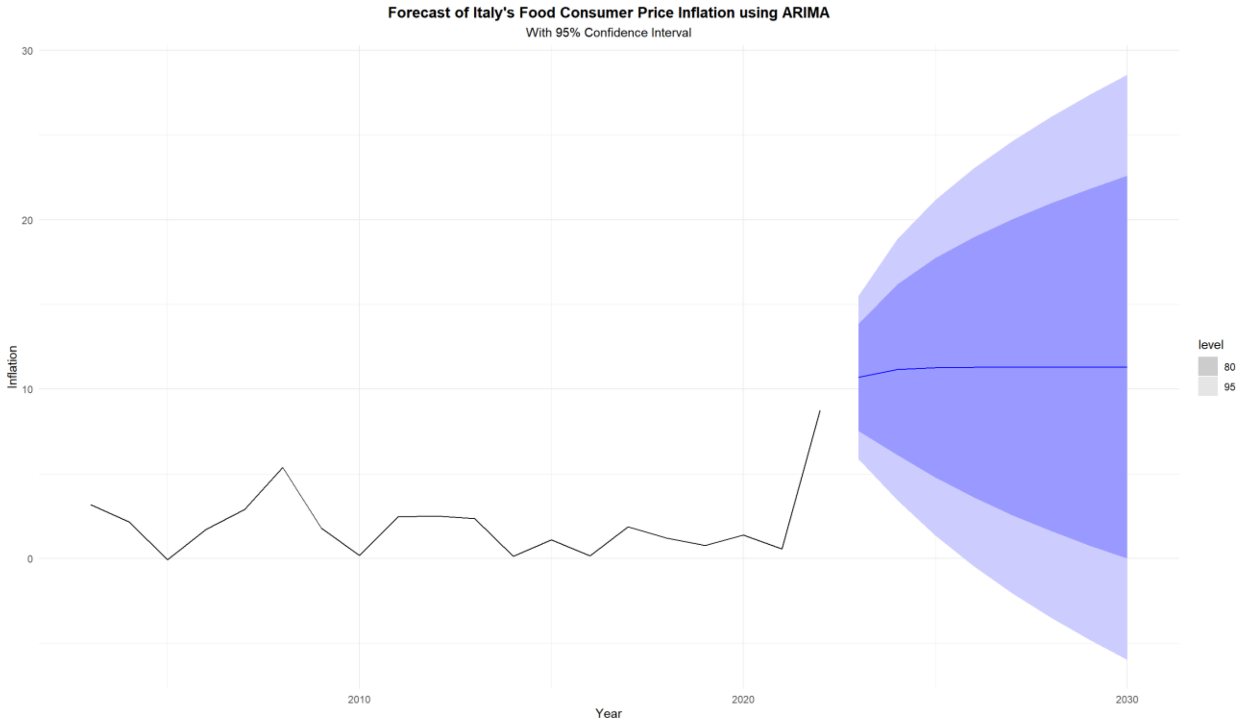

Forecast of Italy's Headline Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
20
10
0
2010
2020
2030
Year
level
80
95

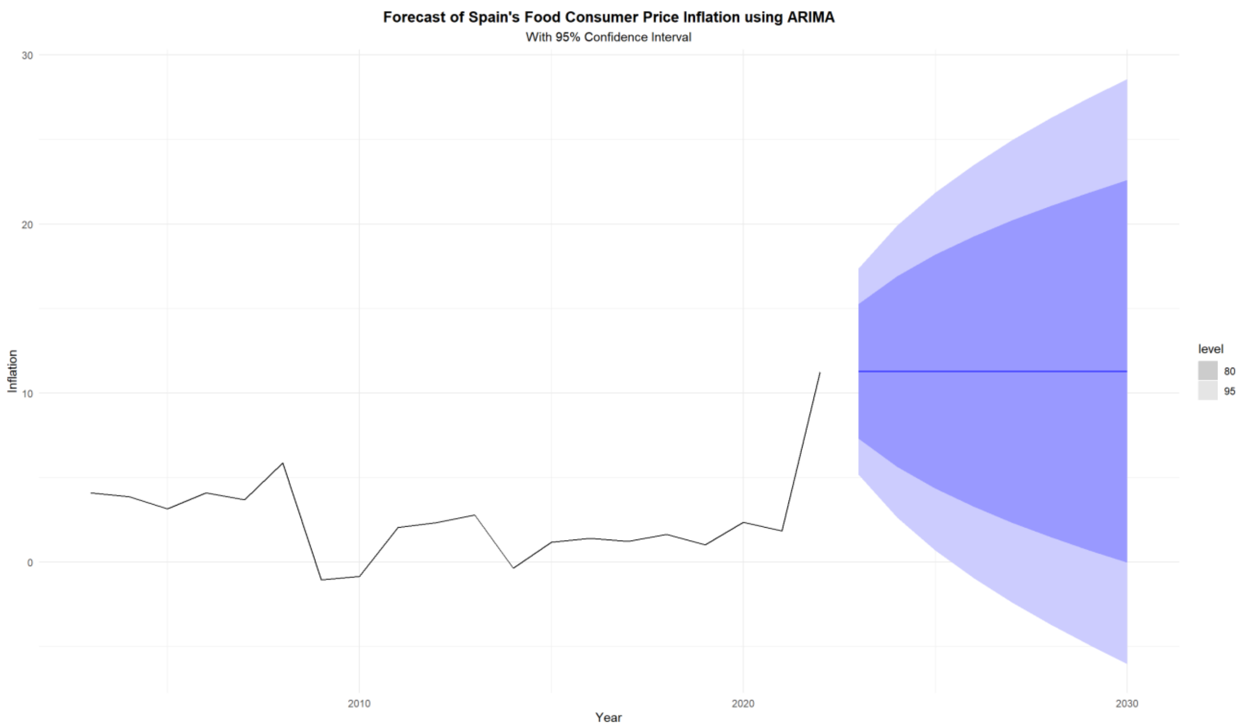

Forecast of Spain's Headline Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
Year
level
80
95
20
15
10
5
0
-5
2010
2020
2030

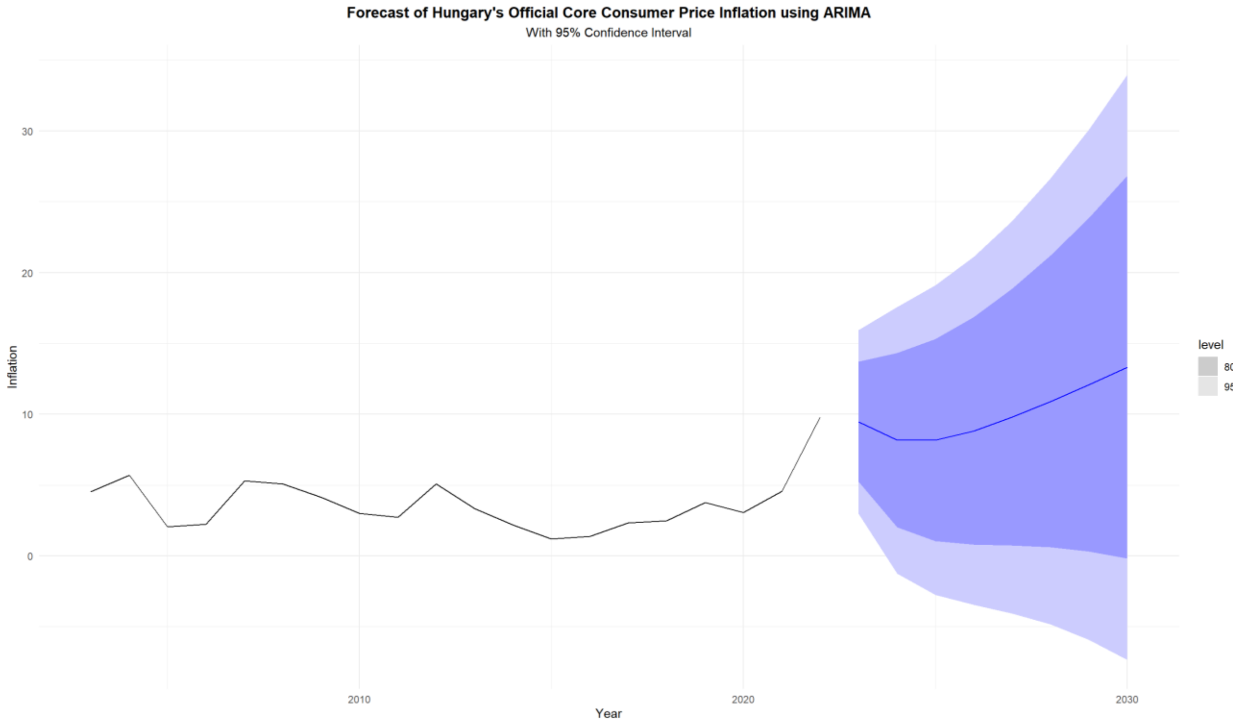

Forecast of Hungary's Energy Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
Year
level
80
95
40
20
0
-20
2010
2020
2030

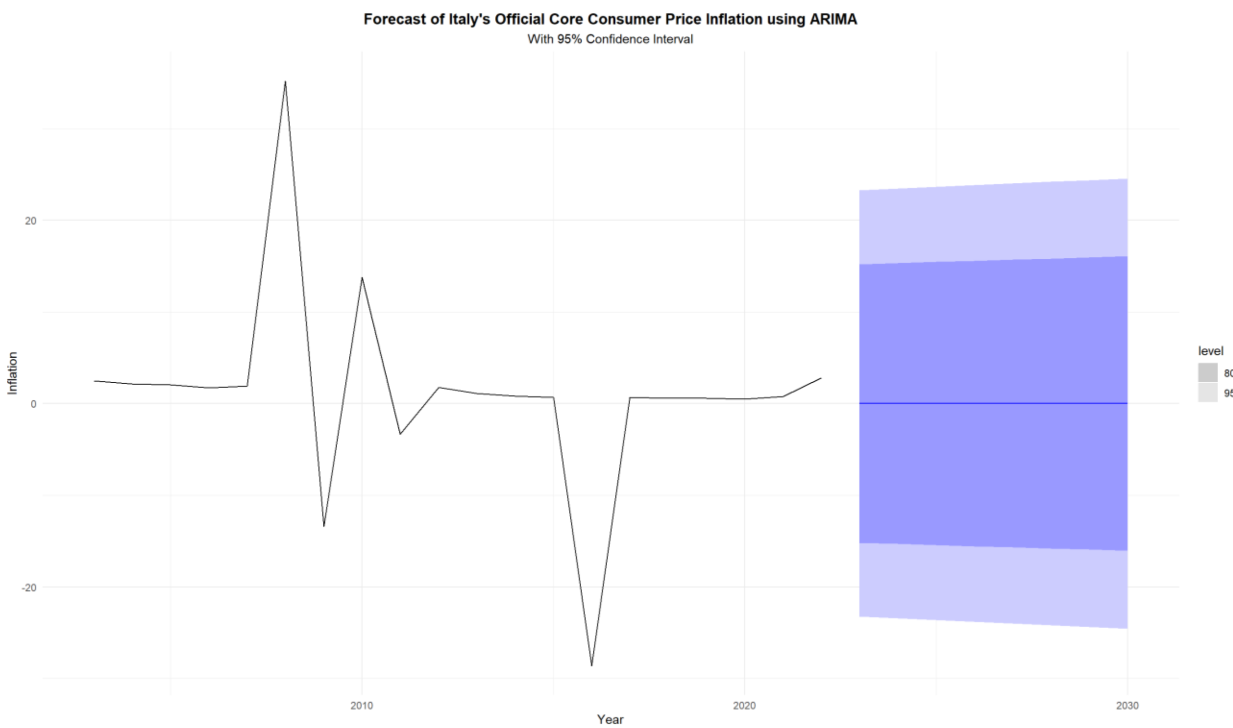

Forecast of Spain's Energy Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
30
20
10
0
-10
2010
2020
2030
Year
level
80
95

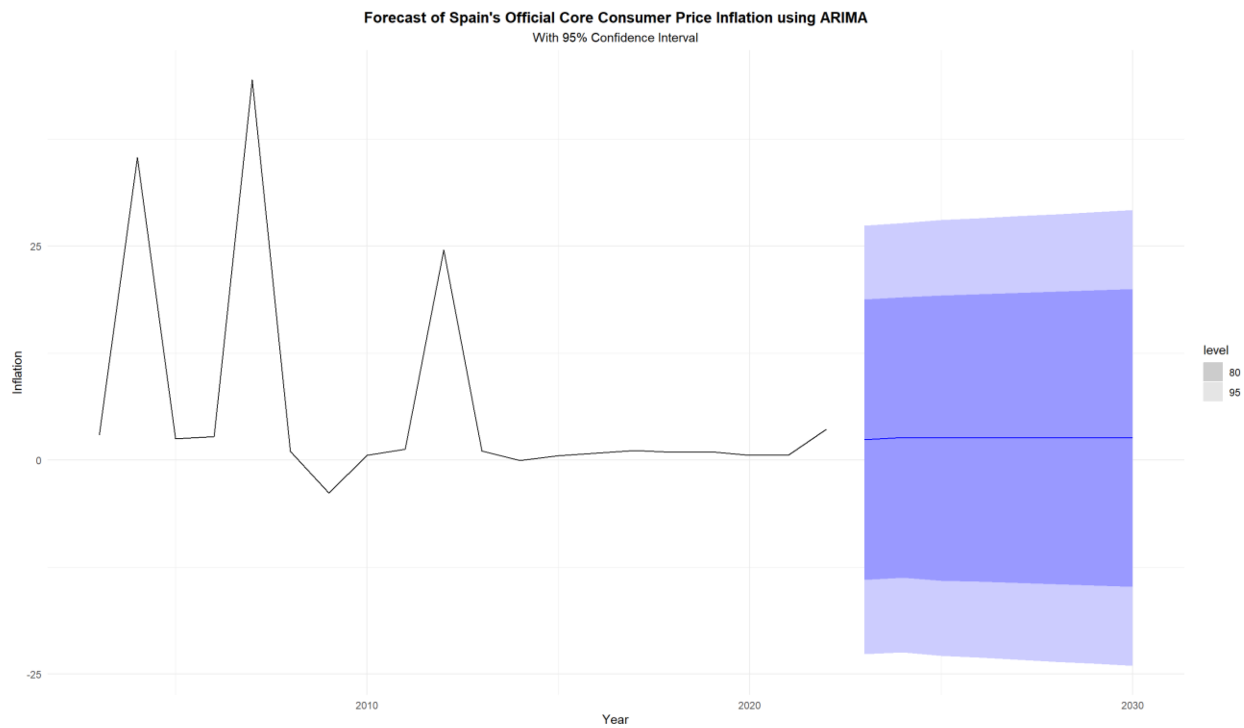

Forecast of Hungary's Food Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
30
20
10
0
-10
2010
2020
2030
Year
level
80
95

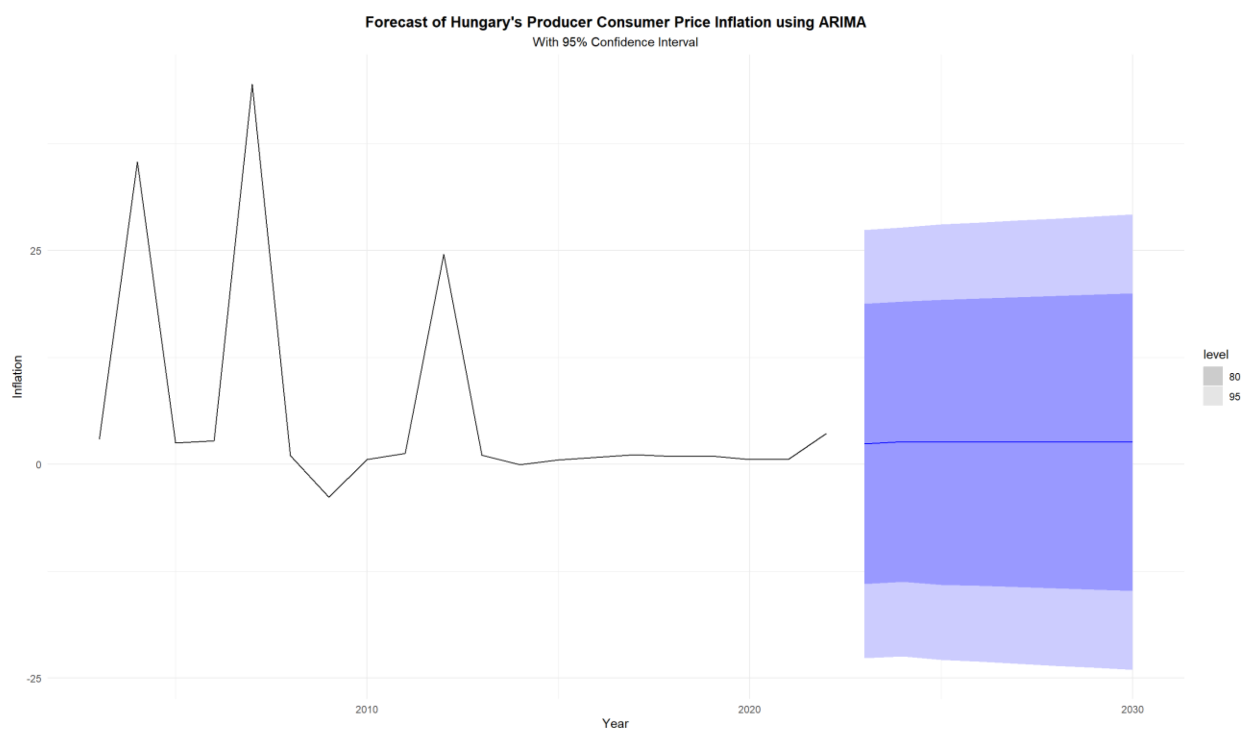

Forecast of Italy's Food Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
Year
level
80
95

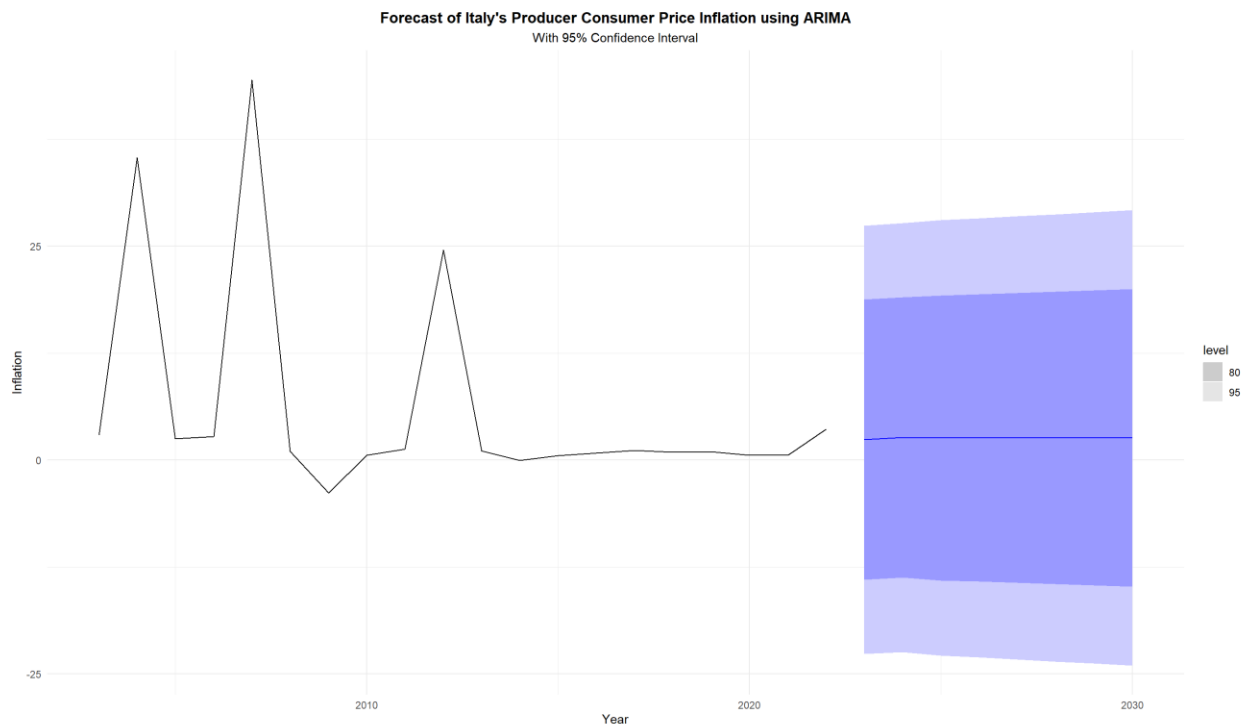

Forecast of Spain's Food Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
Year
level
80
95


Forecast of Hungary's Official Core Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
30
20
10
0
2010
2020
2030
Year
level
80
95

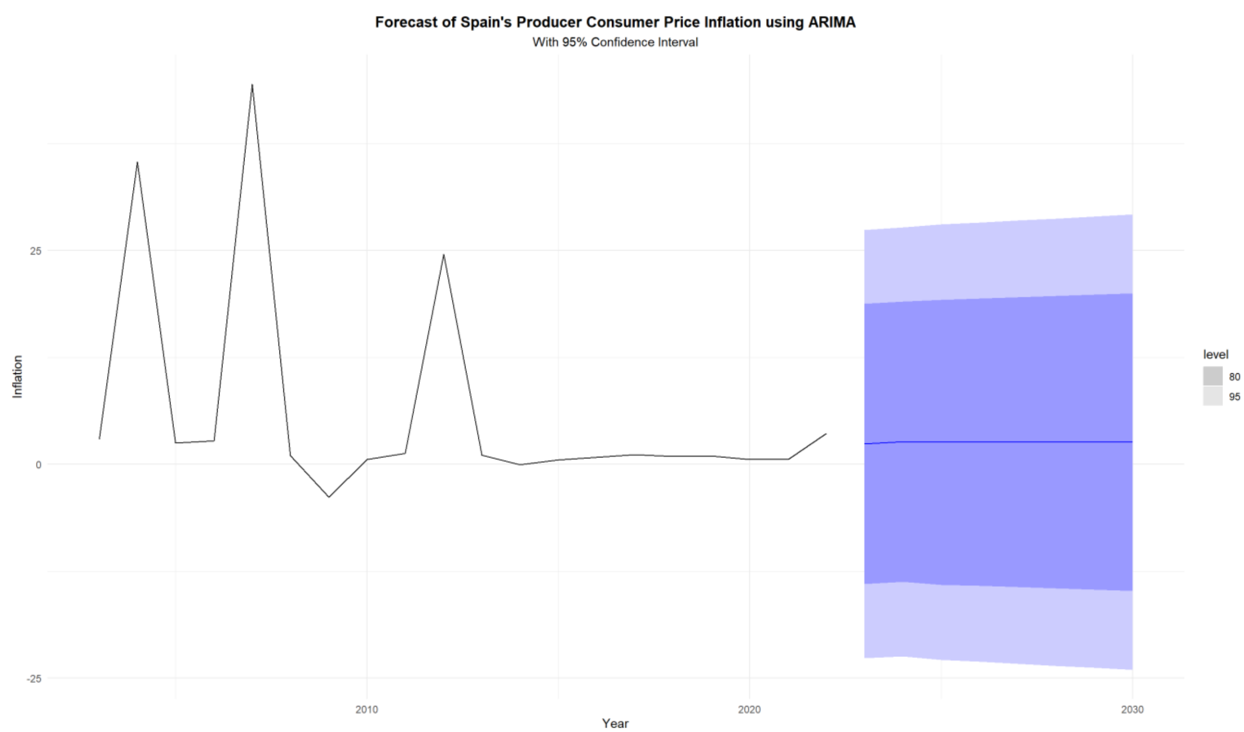

Forecast of Italy's Official Core Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
20
0
-20
2010
2020
2030
Year
level
80
95


Forecast of Spain's Official Core Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
25
0
-25
2010
2020
2030
Year
level
80
95


Forecast of Hungary's Producer Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
25
0
-25
2010
2020
2030
Year
level
80
95



Forecast of Italy's Producer Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
25
0
-25
2010
2020
2030
Year
level
80
95



Forecast of Spain's Producer Consumer Price Inflation using ARIMA
With 95% Confidence Interval
Inflation
25
0
-25
2010
2020
2030
Year
level
80
95