The Italian Harmonised Index of Consumer Prices (HICP): a Time Series analysis (1997-2023)

Foggetti Marco 5111158

Larosa Nicolas 5111178

Table of Contents

[Measuring Inflation - The Harmonised Index of Consumer Prices (HICP) 2](#_Toc127550354)

[Preliminary Analysis 3](#_Toc127550355)

[Removing the influence of Linear Trends on Data 5](#_Toc127550356)

[Removing the influence of Seasonality on Data with Seasonal Differentiation 7](#_Toc127550357)

[Seasonal Differentiation on the First Order Differenced log-transformed Series 9](#_Toc127550358)

[Model Specification 11](#_Toc127550359)

[Model Estimation 13](#_Toc127550360)

[Model Diagnostic 14](#_Toc127550361)

[Analysis of the Variance of the Residuals 14](#_Toc127550362)

[Investigating Residual Normality 15](#_Toc127550363)

[Independence of residuals 17](#_Toc127550364)

[Model Fitting 18](#_Toc127550365)

[Forecasting 19](#_Toc127550366)

[Forecasting from 2023 19](#_Toc127550367)

[Forecasting from 2020 20](#_Toc127550368)

## Measuring Inflation - The Harmonised Index of Consumer Prices (HICP)

In the euro area, the key measure of the rate of **inflation** is the **Harmonized Index of Consumer Prices.** This index is compiled based on data from the national central banks by Eurostat, a Directorate-General of the European Comission responsible for providing statistical information to the European institutions.

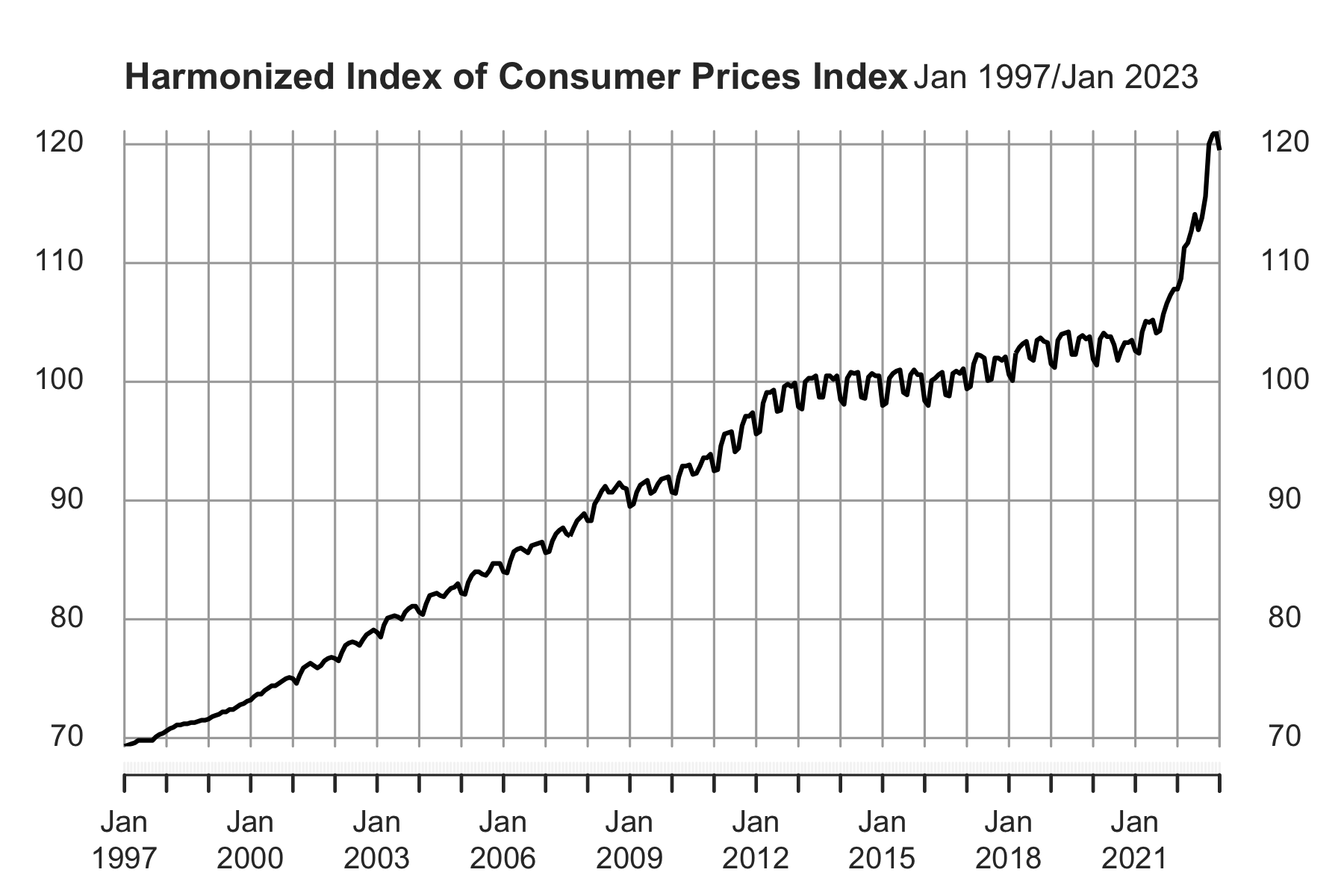
Ensuring *price stability* in the euro area is one of the main responsibilities of the **European Central Bank** (ECB) and the national central banks. ECB aims to keep the annual rate of inflation below but close to 2% in the medium term

This report aims to analyse inflation in **Italy** by analyzing the time series of the harmonised consumer price index, from now on referred to as HICP for brevity.

The aim of this analysis is to fit an ARIMA model, with seasonal components, to HICP. While this model might provide some insides about the inflation rate on its own, it can also be used for forecasting purposes.

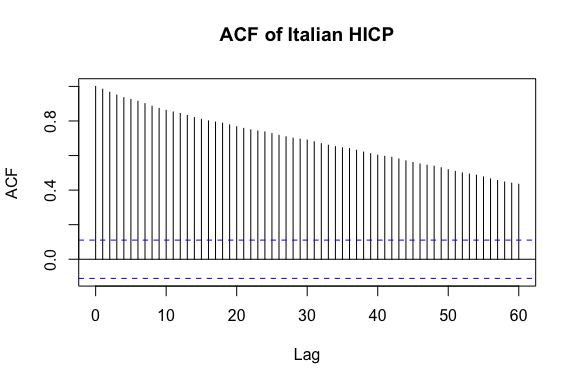
## Preliminary Analysis

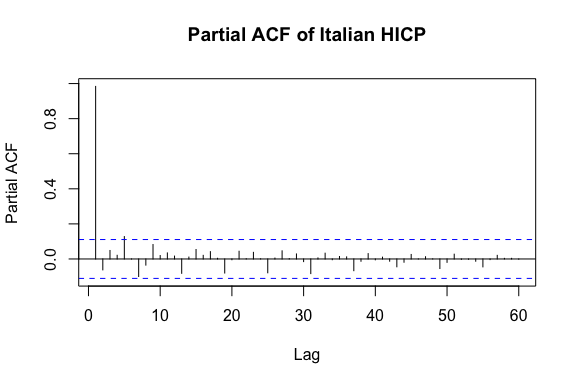
The data comes from the ECB’s “Statistical Data Warehouse” and it is accessed via the R package ‘ecb’ with code ICP.M.IT.N.000000.4.INX. The dataset contains monthly observations about the Italian HICP starting from January 1997 to January 2023 counting 313 values.



Based on the plot above, it is evident that a positive **linear trend** is observed, and as recent observations are approached, there is an increase in variance. Additionally, the presence of **seasonality** can be detected, which is consistent with the characteristics of an economic time series such is the one under analysis.

In order to better understand our data, consider the plot of the autocorrelation function (ACF) and partial autocorrelation function (PACF).

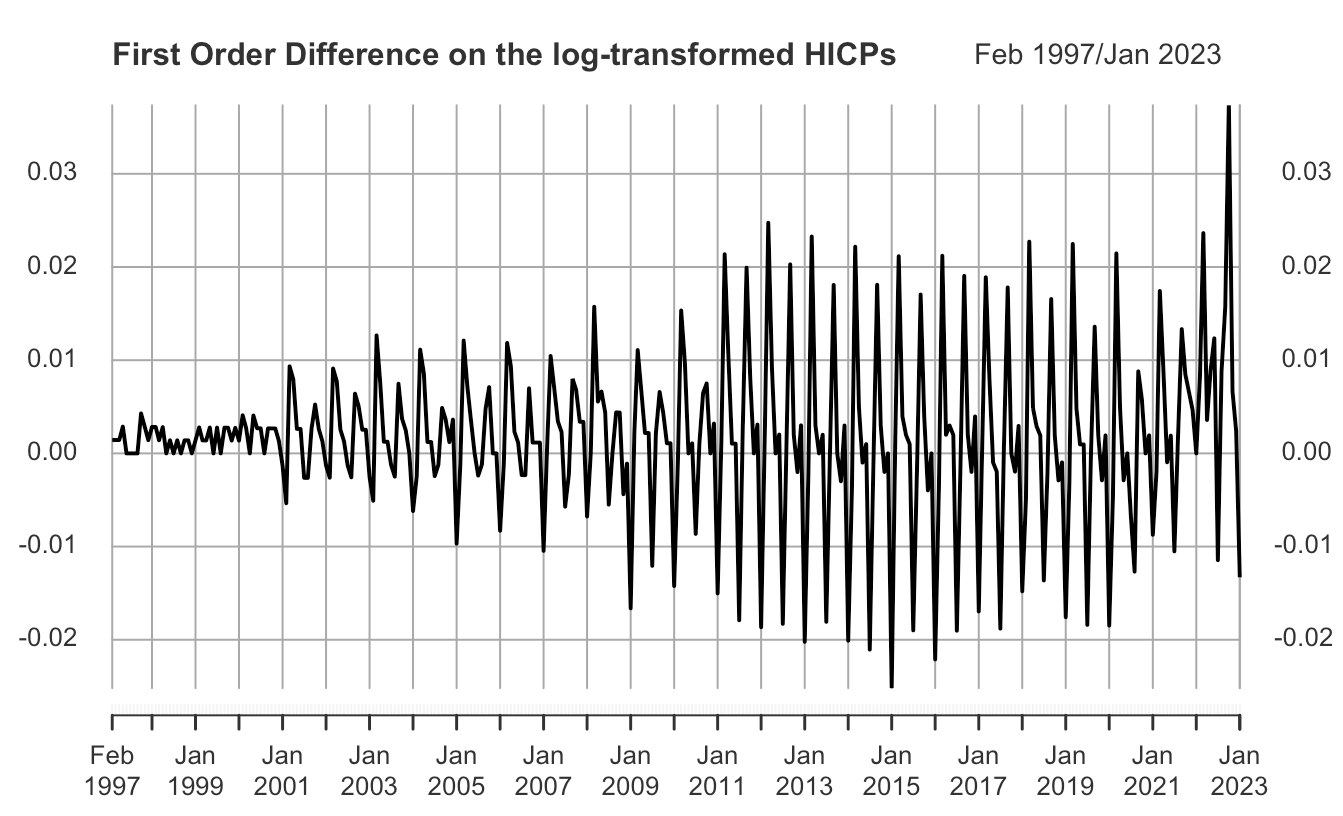




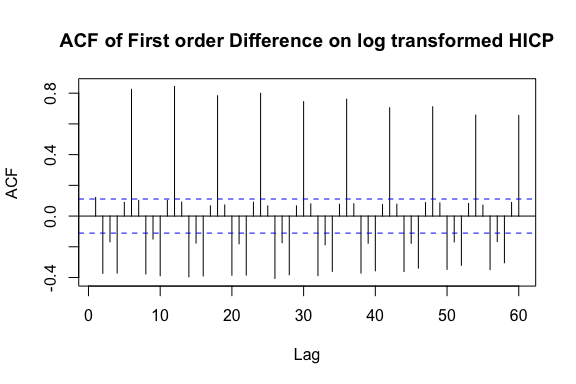
The results of the autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses suggest indeed, as expected, that the time series is characterized by a notable linear trend. Specifically, the ACF plot indicates a gradual decrease in positive autocorrelations over increasing lags, while the PACF plot exhibits a significant positive correlation at lag 1 followed by a sharp decrease at higher lags, which is a characteristic pattern of a time series with a linear trend.

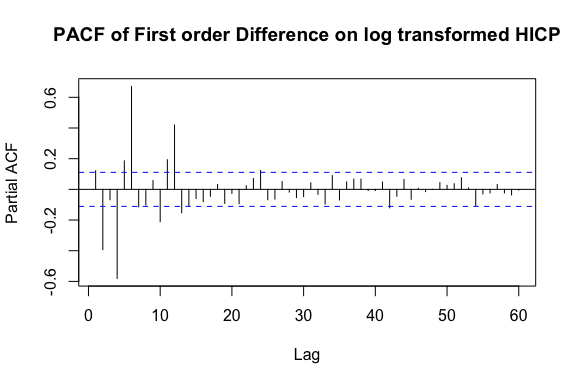
### Removing the influence of Linear Trends on Data

In order to remove the trend and achieve a smoother time series, a **logarithmic transformation** to the data is applied followed by a **first-order difference** on the log-transformed series.



The application of the logarithmic transformation followed by the first-order differentiation on the time series was effective in eliminating the positive trend in the data. Nevertheless, the presence of seasonal patterns may still be present. Therefore, the examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the transformed data will be proceeded with.





The autocorrelation function (ACF) plot provides additional evidence to support the presence of seasonality in the data. The plot shows clear spikes at multiple of lag 6, which are especially pronounced at multiples of lag 12, with a relative slow decay in the autocorrelations. This annual pattern suggest the presence of a seasonal component in the data.

Furthermore, the time series does not exhibit stationarity, as evidenced by the observed increase in variability over time, particularly from around 2013/2014. Therefore, additional transformations are required to achieve stationarity before fitting a SARIMA model. To address the seasonality, it is recommended that a further differentiation with lag 12 has to be considered.

### Removing the influence of Seasonality on Data with Seasonal Differentiation

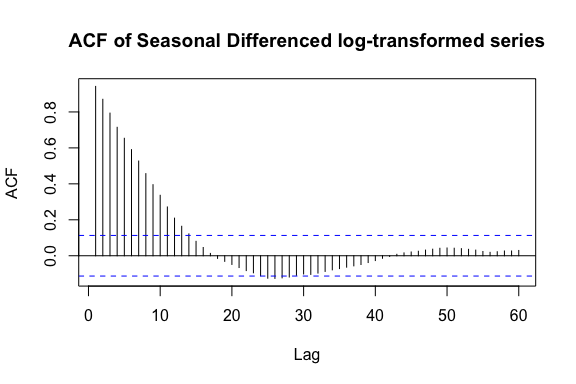
Aiming to remove seasonality on the data a **seasonal difference** with order is applied on the log-trasnformed data.

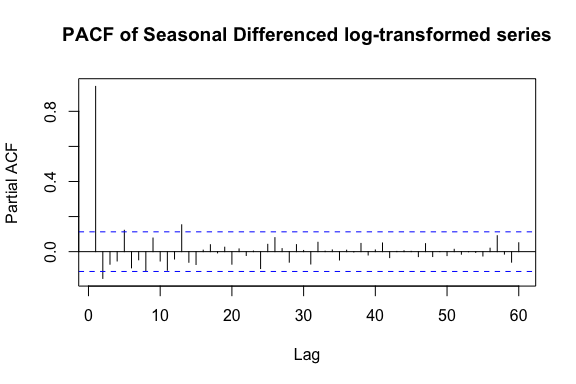
Chart, line chart, histogram

Description automatically generated

Seasonal differencing may have removed the seasonality from the time series. The ACF still shows slow decay indicating some autocorrelation in the data, due to the presence of a linear trend as already mentioned.

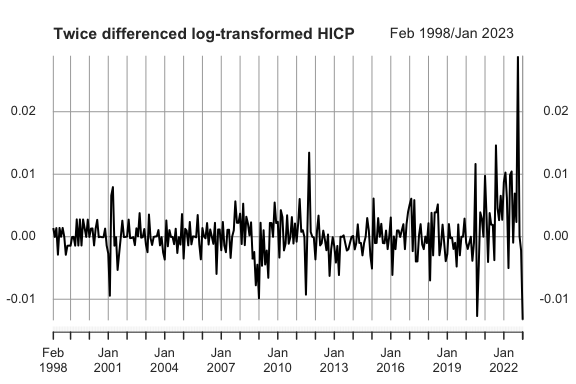
Furthermore, the time series may also exhibit long-term dependence: the slow tail-off of the ACF after seasonal differencing could suggest long-term dependence, meaning that past values of the series are highly correlated with future values, even after accounting for any seasonality in the data, which is due to the presence a trend in the data.



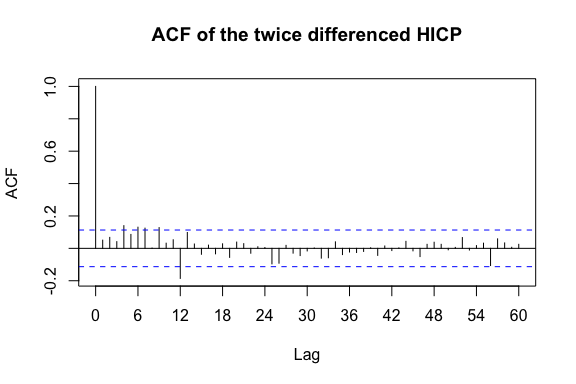


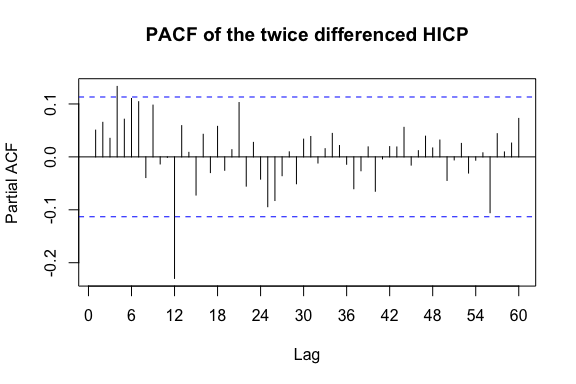
### Seasonal Differentiation on the First Order Differenced log-transformed Series

Hence, to account for both trend and seasonality, the resulting time series is obtained as the log-transformed raw time series twice differenced, first with lag 1 and then lag 12.



The resulting time series now appears to be reasonably stationary so that a sarima model might be fitted.





Just looking at the seasonal lags where it can be noticed that the acf cuts off after a strong peak at indicating that the seasonal part may be modeled with a Seasonal Moving Average process of order 1.

On the other hand, the non seasonal part does not show any particular behaviour. Therefore, the best model will be chosen through criterion model selection using both the AIC and the BIC criteria, assuming moreover the normality of the errors, which will be discussed in the following sections.

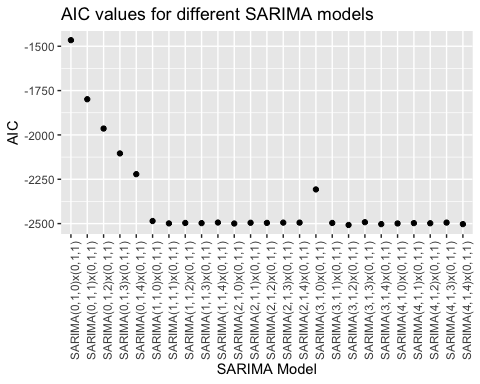
## Model Specification

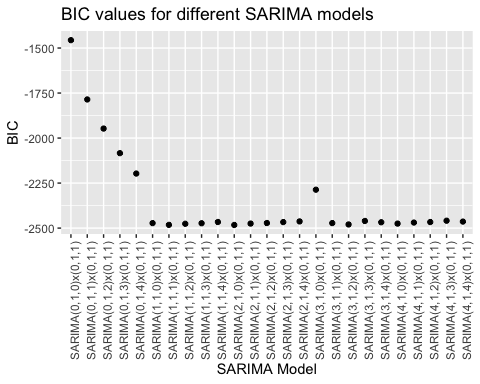
In order to assess the goodness of fit of our models, statistical measures such as the **Akaike Information Criterion** (AIC) and **Bayesian Information Criterion** (BIC) are used. To facilitate the process, a function that computes the AIC and BIC for various model specifications is implemented. Specifically, the seasonal moving average term (SMA) of order is kept constant, while the orders of the non-seasonal component are left varying.

Below are a table and plots summarizing the results, allowing to identify the model with the lowest values of these information criteria as the preferred choice.

|  | AIC | BIC |
| --- | --- | --- |
| SARIMA(0,1,0)x(0,1,1) | -1465.661 | -1456.246 |
| SARIMA(0,1,1)x(0,1,1) | -1798.723 | -1785.601 |
| SARIMA(0,1,2)x(0,1,1) | -1964.035 | -1947.206 |
| SARIMA(0,1,3)x(0,1,1) | -2104.311 | -2083.776 |
| SARIMA(0,1,4)x(0,1,1) | -2221.295 | -2197.052 |
| SARIMA(1,1,0)x(0,1,1) | -2485.471 | -2472.350 |
| SARIMA(1,1,1)x(0,1,1) | -2498.843 | -2482.015 |
| SARIMA(1,1,2)x(0,1,1) | -2496.422 | -2475.887 |
| SARIMA(1,1,3)x(0,1,1) | -2497.321 | -2473.078 |
| SARIMA(1,1,4)x(0,1,1) | -2494.047 | -2466.098 |
| SARIMA(2,1,0)x(0,1,1) | -2499.417 | -2482.588 |
| SARIMA(2,1,1)x(0,1,1) | -2495.337 | -2474.802 |
| SARIMA(2,1,2)x(0,1,1) | -2496.032 | -2471.789 |
| SARIMA(2,1,3)x(0,1,1) | -2494.548 | -2466.599 |
| SARIMA(2,1,4)x(0,1,1) | -2494.551 | -2462.894 |
| SARIMA(3,1,0)x(0,1,1) | -2307.315 | -2286.779 |
| SARIMA(3,1,1)x(0,1,1) | -2496.379 | -2472.137 |
| SARIMA(3,1,2)x(0,1,1) | -2507.715 | -2479.765 |
| SARIMA(3,1,3)x(0,1,1) | -2492.056 | -2460.399 |
| SARIMA(3,1,4)x(0,1,1) | -2502.882 | -2467.518 |
| SARIMA(4,1,0)x(0,1,1) | -2499.201 | -2474.958 |
| SARIMA(4,1,1)x(0,1,1) | -2497.168 | -2469.218 |
| SARIMA(4,1,2)x(0,1,1) | -2498.043 | -2466.386 |
| SARIMA(4,1,3)x(0,1,1) | -2494.338 | -2458.974 |
| SARIMA(4,1,4)x(0,1,1) | -2502.895 | -2463.824 |

|  |  |  |
| --- | --- | --- |
|  |  |  |





Several candidate models possess low values of both AIC and BIC: among them, the model that is the simplest in terms of its structure and parameters, while still adequately fitting the data is:

In terms of Backshift Notation, the model could be expressed as:

where:

* is the backshift operator, such that
* is the autoregressive parameter for the non-seasonal component of the model
* is the moving average parameter for the seasonal component of the model
* is a white noise with zero mean and constant variance
* The term on the left-hand side of the equation represents the differencing of the non-seasonal and seasonal components of the ARIMA model as we are applying a first order difference and a seasonal difference with on the log-transformed series.

## Model Estimation

The ARIMA model is estimated on the mean adjusted log-trasnformed time series data using the *maximum likelihood estimation* method.

The method seeks to maximize the likelihood function, which is the probability of observing the given data for a given set of parameter values.

The obtained estimates of the parameters are the following:

## Model Diagnostic

The assumptions behind the maximum likelihood estimation of a Seasonal ARIMA model are:

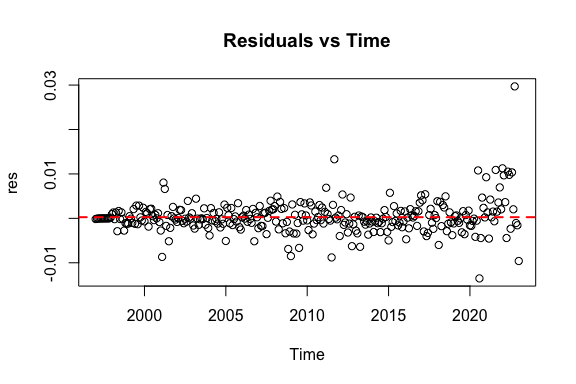
* The errors in the model are **normally distributed**: The maximum likelihood estimation assumes that the errors in the SARIMA model follow a normal distribution. If the errors are not normally distributed, then the maximum likelihood estimates may not be optimal.
* The errors in the model are **independent** and **identically distributed**: The maximum likelihood estimation assumes that the errors in the SARIMA model are independent and identically distributed with mean and variance.

Moreover, If the errors are both uncorrelated and normally distributed, then it can be inferred that they are also independent.

### Analysis of the Variance of the Residuals

When the residuals are identically distributed over time, the statistical model is properly capturing the underlying patterns and trends in the data, and the residuals are not influenced by any particular time points or trends in the time series.

If the residuals are not identically distributed over time, statistical inferences, such as hypothesis tests or confidence intervals, may be biased or incorrect, and predictions made using the model may not be reliable.

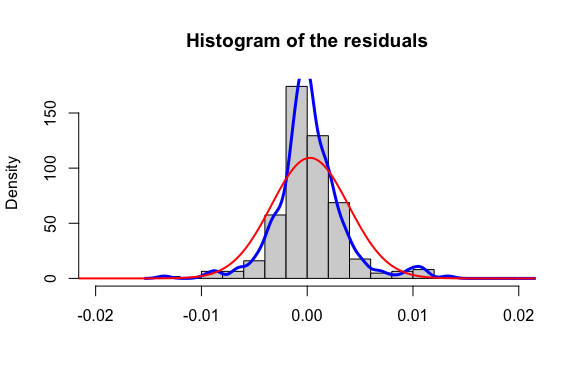


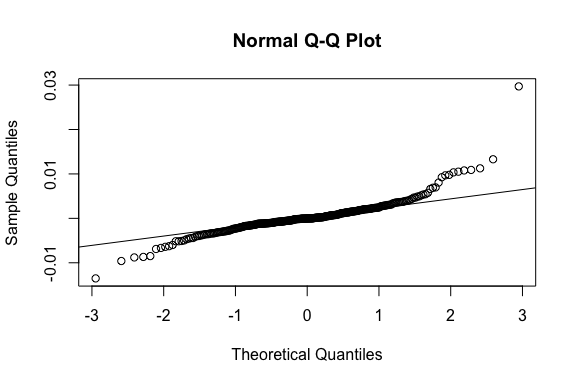
The temporal scatterplot of the residuals displays a rectangular distribution around the horizontal zero line (mean), indicating a lack of discernible pattern in the variability of the residuals with respect to time. As such, there is no evidence of heteroscedasticity in the model with the exception of a few outliers in the years 2020-2023.

The observed peak in inflation during this period can likely be attributed to exogenous factors such as the Covid-19 pandemic and the War between Russia and Ukraine, both of which have had significant impacts on the global economy.

### Investigating Residual Normality

To verify the assumption of normality, several graphical and statistical methods can be employed, including the construction of histograms, QQ-plots, and box-plots, as well as the use of Normality tests such as the Shapiro-Wilk Normality Test.





Both the histogram and the Q-Q plot of the residuals shows departure from normality at the tails due to the presence of the outliers mentioned before.

Overall, the residuals behave well except for the fact that a distribution with heavier tails than normal distribution might be employed to get better results.

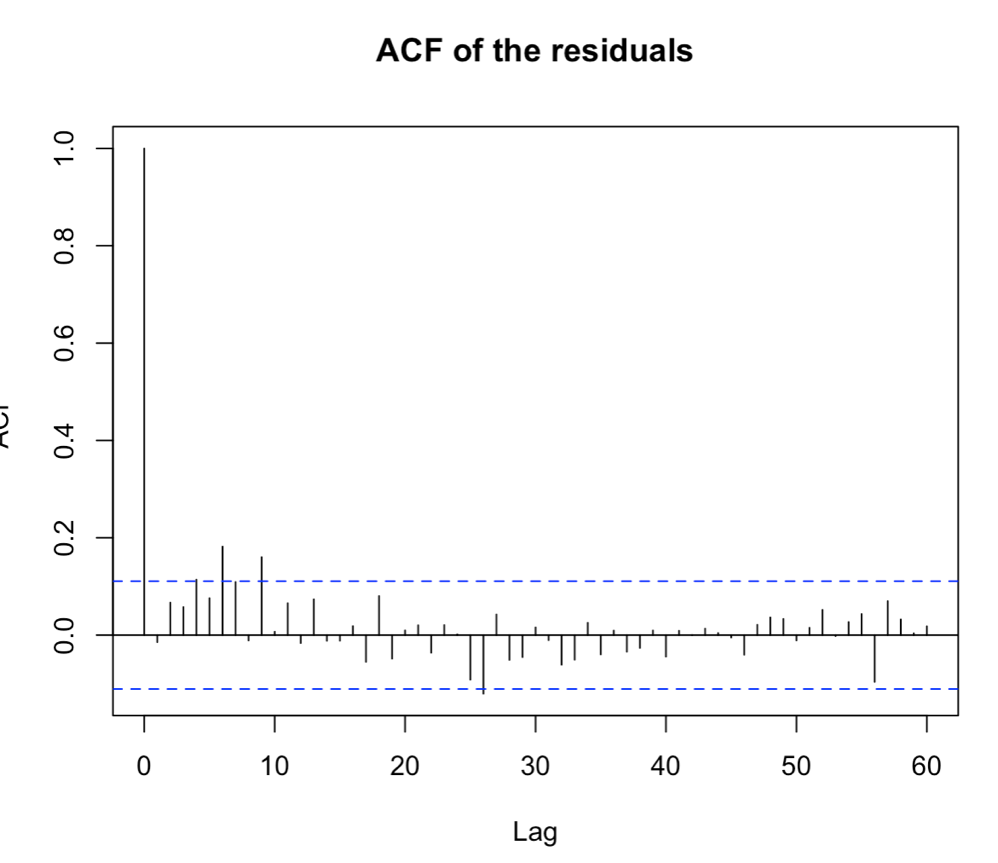
Below we plot the p-value of the Shapiro-Wilk Normality test that, as expected, lead to the conclusion of not accepting the null hypothesis of Normality of errors.

shapiro.test(res)

##   
## Shapiro-Wilk normality test  
##   
## data: res  
## W = 0.85047, p-value < 2.2e-16

### Independence of residuals

In order to check this assumption the Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used.

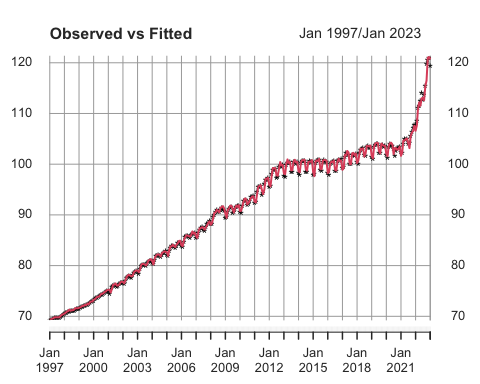


As shown in the plots above, residuals are uncorrelated: the ACF and PACF plots show no significant correlations beyond the first lag except for a small amount of correlation that still remains, although not at seasonal lags.

Chart

Description automatically generated

## Model Fitting



Finally, the visualization of the ‘observed vs fitted’ values suggests that the fitted values provide a good representation of the observed data. Therefore that the model is able to accurately capture the underlying patterns and relationships in the data, and that the model’s predictions closely match the actual data.

Overall, the goodness of fit between the observed and fitted values suggests that the model is reliable for making predictions.

## Forecasting

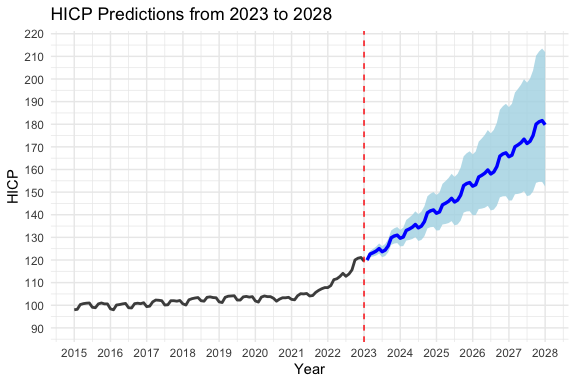
The aim of this section is to make predictions of the level of the HICP for Italy up to 2028 using the estimated SARIMA model.

As already mentioned the last 3 years have been strongly influenced by exogenous factor, causing a sharp increase in the level of inflation.

Thus, it may be worthwhile to consider generating predictions for 2028 using starting points of both 2023 and 2020 to make comparisons.

### Forecasting from 2023

Below is a plot of the predicted values together with the confidence prediction intervals at 95%.



The forecasts exhibit a strong dependency on the previous year’s observations, while adhering to the annual seasonality and a more pronounced linear trend than the trend observed in the series between 1997 and 2020.

This outcome is unsurprising, given that the forecasts are conditional expectations on the observed data and, the forecasting model assigns greater weight to more recent observations, given that the coefficient of the autoregressive part is .

### Forecasting from 2020

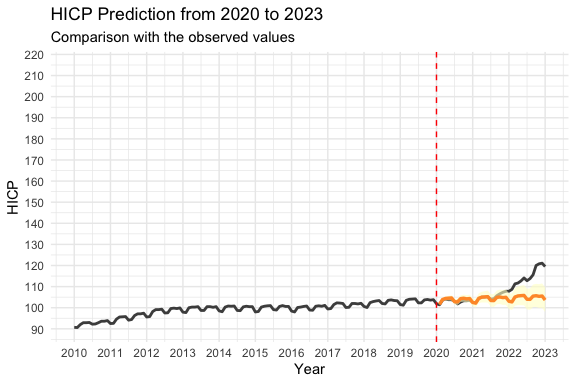
The sarima model is re-estimated without considering the observations from 2020 to 2023.

The new coefficients are now:

#### 

#### Comparison with the Observed Values from 2020 to 2023

As anticipated, the series experienced a significant increase in trend during 2020.



#### 

#### Final comparison

Enhancing the predictions to 2028 the outcome is even clearer: Inflation has experienced a notable increase since 2020 and it may continue to rise.

However, the underlying factors driving inflation are complex and may require a careful analysis of the current economic conditions and trends. Policymakers will need to carefully consider the appropriate measures to mitigate the impact of inflation and support economic recovery in the short and long term.

