American Academic Scientific Research Journal for Engineering, Technology, and Sciences
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402
https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

Could Econometric Models Predict Higher Inflation? Time Series Modelling and Inflation Rate Forecasting

Mir Md. Moheuddin [a]*, Saddam Hossain[b]

[a]*Dept. of Economics and Management, University of Brescia, Brescia-25122, Italy*
[b]*Basic Science Division, World University of Bangladesh (WUB), Dhaka-1205, Bangladesh*
[a]*Email: mirmoheuddin2305@gmail.com*
[b]*Email: s.hossain656@gmail.com*

Abstract

This research utilizes annual time series data on HICP to identify the most suitable econometric model for forecasting the inflation rate in Poland. In this study, we have compared and applied the Autoregressive Moving Average (ARMA) model and the Vector Autoregressive (VAR) model to predict the annual inflation rate using data from January 2002 to December 2020. Various methods are employed to determine the optimal model specifications, followed by the generation of forecasts within a rolling estimation window. The results consistently indicate that the ARMA (2, 0) model outperforms other specifications in forecasting Polish inflation. These findings suggest that the inflation rate is expected to continue its downward trend in the coming years. Consequently, this study offers valuable insights for guiding future actions and policymaking in response to prospective inflation scenarios.

Keywords: Time series data; HCIP; ARMA; VAR; residual analysis; inflation modelling and forecasting.

1. Introduction

Inflation is one of the most crucial economic indicators, employed by governments, stakeholders, politicians, and economists to understand the state of the economy. It is defined as the rate of change in the Consumer Price Index (CPI), a key term that reveals the overall price level of goods and services in a country's economy. We use the Harmonized Consumer Price Index (HICP), which is measured by generally accepted price levels, and it comprises a weighted average of various indexes, including unprocessed food, processed food, non-energy industrial goods, energy, and services, among others. Inflation determines the relative changes in the cost of living as the general price level of goods and services increases.

Received: 8/16/2024
Accepted: 10/16/2024
Published: 10/26/2024

----

* Corresponding author.

Inflation is a vital variable that directly impacts a country's economic situation. Understanding inflation forecasting techniques is crucial for making both fiscal and monetary policies to maintain stability and sustainable economic progress. Controlling inflation at a manageable level is necessary for the balanced development of the economy. Higher inflation can signal economic instability and excessive expenditure, potentially hindering economic development.

Forecasting is a primary objective of econometric modeling, with two main methods: qualitative and quantitative. For central bankers, forecasting is crucial in the decision-making process, as it helps control inflation rates based on economic uncertainty. Making the right decisions regarding forecast values plays a pivotal role in navigating the complex economic system.

This research aims to address central questions at the core of our forecasting analysis: which forecasting method should be utilized? We will compare two classes of methods, namely Univariate (ARMA) models and multivariate (VAR) models, to forecast Polish HICP inflation.

The primary objective of this thesis is to model and forecast inflation rates in Poland as accurately as possible. We will compare two forecasting methods: the Autoregressive Moving Average (ARMA) model and the Vector Autoregressive (VAR) model, for predicting Polish HICP inflation as a rate of change. VAR is a standard tool used for long-term forecasting, while ARMA is relatively robust for short-term forecasting. We will use these models to forecast the yearly inflation rate and determine the highest forecast accuracy using a range of criteria to identify optimal forecasts in our study.

2. Literature review

In the current era, statistical and economic modelling are widely utilized in empirical and theoretical research. Quality prediction is essential for helping investors understand market behaviors and improve their profitability based on applied modeling techniques. Policymakers want to know about the pattern of inflation rates and modelling is also helpful for them to apply appropriate policies to control inflation. Nowadays, these types of research play a significant role in developing economic sectors as well as leading the transparent relationship between variables and external factors.Khan, S., & Alghulaiakh, H. [1] applied and compared auto ARIMA (Auto Regressive Integrated Moving Average model). They showed two customized ARIMA (p, D, q) to get an accurate stock forecasting model by using Netflix stock historical data for five years. Szafranek K. [2] utilized a thick modelling approach to investigate the quality of the out-of-sample short-term headline inflation forecasts generated by a combination of bagged single hidden-layer feed-forward artificial neural networks. Facebook’s Prophet Forecasting Model and ARIMA Forecasting Model were utilized to compare their performance and accuracy on a dataset containing the confirmed cases, deaths, and recovered numbers, obtained from the Kaggle website (Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H.,). In their paper, the forecast models were then compared to the last 2 weeks of the actual data to measure their performance against each other, and the result showed that Prophet generally outperforms ARIMA, despite it being further from the actual data the more days it forecasts [3].

Abonazel, M. R. and his colleagues, [4] used the Box-Jenkins approach to build the appropriate Autoregressive-Integrated Moving-Average (ARIMA) model for the Egyptian GDP data. Egypt's annual GDP data was obtained from the World Bank for the years 1965 to 2016. Khan, Firdos, Alia Saeed, and Shaukat Ali [5] demonstrated modelling and forecasting of new cases, deaths and recovery cases of COVID-19 by using the Vector Autoregressive model and their forecasted model results showed a maximum of 5,363/day new cases with a 95% confidence interval of 3,013–8,385 on 3rd of July, 167/day deaths with 95% confidence interval of 112–233 and maximum recoveries 4,016/day with 95% confidence interval of 2,182–6,405 in the next 10 days. Ramyar, S. and his colleagues, [6] tried to develop a multilayer perceptron neural network and trained historical data from 1980 to 2014 and using mean square error for testing data, the optimal number of hidden layer neurons was determined and the designed MLP neural network was used for estimation of the forecasting model. The exchange rate played a dominant role as a policy instrument in the 1990s (Gottschalk, J. and Moore, D.). In their papers, linkages between the short-term interest rate and inflation had been weak [7]. Moser, G. and his colleagues, [8] applied factor models proposed by Stock and Watson and VAR and ARIMA models to generate 12-month out-of-sample forecasts of Austrian HICP inflation and its subindices processed food, unprocessed food, energy, industrial goods, and services price inflation. Rumler, F., Moser, G. and Scharler, J. Reference [9] evaluated the performance of VAR and ARIMA models to forecast Austrian HICP inflation. Additionally, they investigated whether disaggregate modelling of five subcomponents of inflation is superior to specifications of headline HICP inflation.

Uko, A. K and Nkoro, E [10] examined the relative predictive power of ARIMA, VAR and ECM models in forecasting inflation in Nigeria. Comparatively, they examined the performance of the forecasting ability of the models, and how well the simulated series track the actual data. Poulos and his colleagues, [11] described an automated forecasting system that encompasses an objective ARIMA method with the Holt-Winters procedure in a weighted averaging scheme. Arratibel, O. and his colleagues, [12] provided stylized facts on monetary versus non-monetary (economic and fiscal) determinants of inflation in these countries as well as formal econometric evidence on the forecast performance of a large set of monetary and non-monetary indicators. Hubrich K. [13] analyzed whether the accuracy of forecasts of aggregate euro area inflation can be improved by aggregating forecasts of subindices of the Harmonized Index of Consumer Prices (HICP) as opposed to forecasting the aggregate HICP directly. The analysis included univariate and multivariate linear time series models and distinguished between different forecast horizons, HICP components and inflation measures. Pufnik A and Kunovac D. [14] attempted to forecast changes in the index's components to obtain a more detailed insight into the sources of future inflationary or deflationary pressures and to determine whether a forecast of developments in the total consumer price index obtained by aggregating forecasted values of the index's components is more precise than a direct forecast. Junttila J. [15] forecasted the rate of future inflation in Finland for the time period of unregulated financial markets since the beginning of 1987. Wigati, Y., Rais, R., & Utami, I. T. [16] developed a model with the best time series Autoregressive Integrated Moving Average (ARIMA) to predict the movement of data Consumer Price Index (CPI) in Palu - Central Sulawesi.

Mohamed J. [17] compared Autoregressive Integrated Moving Average (ARIMA) and regression with ARIMA errors, where the covariate is the time, to forecast Somaliland Consumer Price Index using monthly time series data from 2013 – 2020. Tandon H. and his colleagues. [18] developed for forecasting future COVID-19 cases in

India. In their papers, the time series analysis indicated that the cases would keep on increasing in India in the coming month as the peak. Djawoto, D. [19] researched to forecast the inflation rate in November 2010 with the Consumer Price Index (CPI) by using ARIMA. He also mentioned in his analysis that the ARIMA method is sufficient for short-term forecasting, whereas if used for long-term forecasting, the resulting value will tend to be constant. Kumar M and Anand M. [20] had been used a time series modelling approach (Box-Jenkins' ARIMA model) in their study to forecast sugarcane production in India.

3. Data description

The dataset used in this study is secondary data based on the Harmonised Consumer Price Index (HICP). This quantitative dataset covers the period from January 2002 to December 2020 and is published by the European Statistical System (ESS) online on the official website of the European Union (https://ec.europa.eu/eurostat). For the purpose of our research, we analyzed monthly HICP data to forecast Polish HICP using R software (version 3.6.3). We applied both the Autoregressive Moving Average (ARMA) model and the Vector Autoregressive (VAR) model using the R programming language in R Studio.

3.1 Autoregressive Moving Average (ARMA) model

3.1.1 Introduction of ARMA

The econometric models include several popular time series models and model testing processes like simple autoregressive (AR) models, simple moving average (MA) models, mixed autoregressive moving-average (ARMA) models, unit-root nonstationary, regression models with time series errors, etc. The ARMA model comprises two processes: the Autoregressive (AR) model, which establishes a relationship between past and present values, and the Moving Average (MA) model, which indicates that the present value depends on past errors. For practical representation, it is desired to obtain models which drive parameters parsimoniously. The parsimony has been gained through a linear process with a small number of parameters for the autoregressive-moving average (ARMA) model. So, the ARMA model takes into consideration a parsimonious explanation of a (weakly) stationary stochastic process for two polynomials, one for the AR model and another for the MA model.

In our study, we used the Autoregressive Moving Average model (ARMA) for analyzing univariate time series and forecasting. With the ARMA model, researchers can easily understand equilibrium and how inflation reverts to equilibrium. It also helps in analyzing the impact of shocks on our series and how inflation reacts to such shocks.

3.1.2 Autoregressive (AR) model

The autoregressive model is the response variable as a function of past values (i.e., lag of the series). For a series y_t, an autoregressive process of order p, AR(p) can be written as:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots \dots \dots + \alpha_p y_{t-p} + u_t; \;\; t = 1,2, \dots \dots T \qquad (1)$$

Alternatively, we can write $y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + u_t$.

Where, y_t is the response variable at time t, α_i,…., α_p are the parameters of the model, y_{t-i} is the dependent variable at time t - i (i =1, 2, …., p), the random variable u_t is the error term of white noise process at time t assumed independently and identically distributed (iid) normal random variables with $E(u_t) = 0$ and $var(u_t) = \sigma^2$; i.e. $u_t \sim iid\ N(0, \sigma^2)$. Using the backshift operator B, the autoregressive (AR) model can be defined in the equivalent form,

$$(1 - \alpha_1 B - \alpha_1 B^2 - \cdots .. - \alpha_p B^p) y_t = u_t$$

Alternatively, we can define, $\alpha(B) y_t = u_t$

Besides, all preceding values of this model may have an additive impact on the series of y_t and so on. So, this model is also called a long-term memory model.

3.1.3 Moving-average (MA) model

This model is a linear form of a series model that is dependent on past errors. For a time-series y_t, a moving average process of order q, MA(q) can be expressed as follows.

$$y_t = u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots \ldots .. - \beta_q u_{t-q};\ t = 1, 2, \ldots \ldots T \qquad (2)$$

Or we can write $y_t = u_t - \sum_{i=1}^{q} \beta_i u_{t-i}$.

Whereas $\beta_1, \beta_2, \ldots \ldots, \beta_q$ are model parameters and q is the lags of this MA model.

The moving-average model is considered as the past errors of explanatory variables. Hence, only q errors can affect the dependent variable (y_t), but larger order errors do not affect on y_t. That means it is a short-term memory model. Applying the backshift operator Bu_t= u_{t-1}, the moving-average (MA) model can define in the equivalent form as

$$y_t = (1 - \beta_1 B - \beta_1 B^2 - \cdots .. - \beta_q B^q)\, u_t$$

Or more succinctly as $y_t = \beta(B) u_t$

Now, adding equations (1) and (2), we can get ARMA (p, q) i.e., the combination of the AR(p) and MA (q) model which constructs the ARMA (p, q) model is written below,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots \ldots \ldots + \alpha_p y_{t-p} + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \cdots \ldots .. - \beta_q u_{t-q}$$

Alternatively, $y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + u_t - \sum_{i=1}^{q} \beta_i u_{t-i}$

Where, y_t and u_t are the actual value and random error at time t; $\alpha_m(m = 1,2,3, \ldots \ldots, p)$ and $\beta_n(n = 1,2,3, \ldots \ldots, q)$ are model' parameters, p and q are considered as integers for the order of AR (p) and MA (q), the error term u_t is assumed to be independent and identically distributed (i.i.d) with a mean value of zero and constant variance σ^2. Again, using the backward shift operator, the ARMA (p, q) can be expressed in the following form,

$$\alpha(B)y_t = \beta(B)u_t$$

Where, $\alpha(B) = 1 - \alpha_1 B - \alpha_1 B^2 - \cdots .. - \alpha_p B^p$ and $\beta(B) = 1 - \beta_1 B - \beta_1 B^2 - \cdots .. - \beta_q B^q$

3.1.4 Identification of correlograms

To assess graphical stationary and determining the appropriate values for the order p in AR(p) and the order q in MA(q), the formal test procedure is applied using the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF measures the correlation between observations in a time series for a set of lags. It identifies which lags have significant correlations, helping us understand the patterns and properties of the time series and enabling us to determine the order of an MA model. The basic definition of ACF is given below,

The autocorrelation function (ACF) at lag k, denoted as s_k, of a stationary stochastic process is defined as

$$s_k = \frac{\rho_k}{\rho_0}$$

Where, $\rho_k = cov(y_t, y_{t+k})$ for all t and ρ_0 is the variance of the stochastic process.

Mathematically, the mean of a series $y_1, \ldots, y_T$ is in the following,

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_i$$

The auto covariance function at lag k, for k ≥ 0, of the time series is defined by

$$\rho_k = \frac{1}{T}\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y}) = \frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})$$

Now, the autocorrelation function (ACF) at lag k, for k ≥ 0, of the time series is defined by,

$$s_k = \frac{\rho_k}{\rho_0} = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

Where, ρ_0 is the variance of the time series and the plot of S_k with lag, k is called a correlogram.

The PACF tells us about the partial correlation between the series and its lag values. It shows the correlation between the observation at the current time and at the preceding time spots and finds out the order of an AR model. The theoretical PACF is in the following form,

$$r_{k,k} = \frac{s_k - \sum_{i=1}^{k-1} r_{k-1,i}\, s(k-i)}{1 - \sum_{i=1}^{k-1} r_{k-1,i}\, s(i)}$$

Where, $r_{k,i} = r_{k-1,i} - r_{k,k} r_{k-1,k-i}$ for i = 1, 2, …., k-1 and s_k is the autocorrelation function.

3.1.5 Unit root (UR) test

This test is conducted to check whether the time series data is stationary or not. The Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test were utilized for the UR test. The test is considered on the assumption that a time series data y_t follows a random walk:

$$y_t = \sigma y_{t-1} + u_t$$

Where, $\sigma = 1$ and y_{t-1} will be subtracted from both sides. We get, $\Delta y_t = \delta y_{t-1} + u_t$ and $\delta = \sigma - 1$. The null hypothesis is H0: $\delta = 0$, i.e., the series is non-stationary and therefore the alternative hypothesis is H1: $\delta < 0$, i.e., the series is stationary and $\sigma < 1$.

3.1.6 Model specification

We use information criteria to select the best models in terms of the in-sample fit.

AIC: Akaike information criterion (Akaike, 1974),

AIC = $-2\frac{l}{T} + 2\frac{K}{T}$

SIC: Schwarz information criterion is known as Bayesian information criteria (Schwarz, 1997),

BIC = $-2\frac{l}{T} + 2\frac{K}{T}\ln(T)$

HQIC: Hannan-Quinn information criterion (Hanna and Quinn, 1979),

HQIC = $-2\frac{l}{T} + 2\frac{K}{T}\ln(lnT)$

Where K is the number of estimated parameters and $l = \ln(\psi)$ is the log-likelihood. Generally, the BIC penalizes free parameters stronger than the AIC and good models are gained by decreasing the value of AIC, HQIC, and BIC and optimizing the log-likelihood. We choose the best model with the lowest information criteria (IC). Note that: K(BIC) ≤ K(HQIC) ≤ K(AIC), IC depends on the fit (log-likelihood) and penalty on the number of parameters.

3.1.7 Likelihood ratio (LR) test

In this test, we can compare the model fit.

H0: the fit of the big ARMA (m parameters more) is the same as the fit of the small ARMA model.

LR = $-2(l_r - l_u) \sim \chi^2(m)$

Where m is the number of additional parameters, and l is the log-likelihood for restricted (small) and unrestricted (big) models.

3.1.8 Model validation and forecasting using the ARMA model

The estimated model will be considered a good model if it simulates the historical behavior properly. The quality of the residual term is assessed through diagnostic tests based on residuals. The Ljung-Box test is utilized to verify the overall adequacy of the estimated model. One notable feature of the ARMA model is its ability to make accurate forecasts. To predict the future values of the time series, we use the given point forecast equation (calculated recursively),

$$y_{T+h}^{f} = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{T+h-i}^{f} + u_t - \sum_{i=1}^{q} \beta_i u_{T+h-i}^{f}$$

Where, $u_{T+h}^{f} = 0$ for h > 0 and $y_{T+h}^{f} = y_{T+h}$ for h ≤ 0.

The forecast error is:

$$y_{T+h} - y_{T+h}^{f} = \sum_{h=0}^{H-1} \theta_h \, u_{T+H-h}$$

Also, the forecast variance (only due to stochastic term) is:

Var $(y_{T+h}) = \sum_{h=0}^{H-1} \theta_h^2$

3.2 The Vector Autoregressive (VAR) model

3.2.1 Introduction of VAR

VAR is one of the most successful, flexible models applied for multivariate time series data. The VAR model establishes dynamic relationships within the structural model without limitations, where all variables are jointly considered endogenous. The VAR model provides a way to the structural form of higher-scale simultaneous equations for which it is differed from the system of simultaneous equations. In the context of the VAR model, the availability of lagged values of dependent variables enables it to make robust predictions about future economic developments. This model's purpose is not solely to assess one-way relationships among variables but also to reveal linkages between variables through lagged effects.

It is a valuable tool extensively used to express the dynamic behavior of macro econometric modeling, financial time series analysis, and forecasting. So, the model is an n-variables, n-equations model, which expresses each variable as a linear function of its past values, the past values of all other variables being treated, and a serially uncorrelated error term. VAR models maintain a coherent and credible approach to describing data, inferring structural relationships, and conducting policy analysis.

3.2.2 The Vector Autoregressive (VAR) model

For a vector of time series data y_t, a general VAR (p) model with n endogenous variables, p lags, and m exogenous variables can be expressed mathematically in the following form

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots \ldots + A_p y_{t-p} + B x_t + u_t, u_t \sim \mathcal{N}(0, \Sigma), t = 1, 2, \ldots, T$$

Where, $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{n,t})$ is an $\mathrm{n} \times 1$ vector of endogenous data, $A_1, A_2, \ldots., A_p$ are p matrices of parameters in dimension $n \times n$, B is a $n \times m$ matrix, and x_t is a $m \times 1$ vector of exogenous regressors that can be constants terms, time trends, or exogenous data series, $u_t = (u_{1,t}, u_{2,t}, \ldots., u_{n,t})$ is a vector of residuals (serially uncorrelated or independent) with a time-invariant covariance matrix Σ .

The simplest VAR (p) model with two variables and lag p =1, can be written in matrix form (more compact notation) as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$

Also, we can write the two systems of equations,

$$y_{1,t} = a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + u_{1,t}$$

$$y_{2,t} = a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + u_{2,t}$$

3.2.3 Estimating VAR model

In compact form, the VAR model can be written (Source: Dieppe, 2016),

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots \ldots + A_p y_{t-p} + B x_t + u_t, where\ t = 1, 2, \ldots, T \qquad (1)$$

Where, $u_t = (u_{1,t}, u_{2,t}, \ldots., u_{n,t})$ is a vector of residuals following a multivariate normal distribution.

$$u_t \sim \mathcal{N}(0, \Sigma) \qquad (2)$$

For further calculations, a reformulation of (1) consists of a transpose form as

$$y_t{}' = y'_{t-1}A'_1 + y'_{t-2}A'_2 + \cdots \ldots + y'_{t-p}A'_p + x'_tB' + u'_t, where\ t = 1, 2, \ldots, T$$

Reformulating the model into a single matrix, we get VAR(p) in matrix notation,

$$Y = XC + \varepsilon \qquad (3)$$

An estimate $\hat{C}$ (LS estimates) of the parameter C in (3) obtains from:

$$\hat{C} = (X, X)^{-1}X, Y \qquad (4)$$

An OLS estimate $\hat{\varepsilon}$ (residuals) of the residual matrix ε can be obtained from the direct application of (3):

$$\hat{\varepsilon} = Y - X\hat{C}$$

Also, a (degree of freedom-adjusted) estimates $\hat{\sum}$ of the covariance matrix $\sum$ in (2) can be gained from:

$$\hat{\sum} = \frac{1}{T-k-1}(\hat{\varepsilon}\ ,\ \hat{\varepsilon})$$

Where k = 1+np is the number of parameters in each equation.

3.2.4 Model validation and forecasting

When the model's parameters are estimated, it is essential to perform diagnostic checks to assess the quality of the time series model. After investigating and identifying the data-generating process for time series models, it becomes useful for forecasting h steps ahead. Generally, the point forecast for horizon h is given below,

$$y_{t+h|t} = A_1y_{t+h-1|t} + A_2y_{t+h-2|t} + \cdots \ldots + A_py_{t+h-p|t}$$

The VAR moving average (VMA) representation is given as

$$y_t = \phi_0u_t + \phi_1u_{t-1} + \phi_2u_{t-2} + \cdots ..$$

The error of forecast for horizon h due to future shocks,

$$y_{t+h} - y_{t+h|t} = \phi_0u_{t+h} + \phi_1u_{t+h-1} + \cdots .. + \phi_{h-1}u_{t+1}$$

The variance of forecast error for horizon h due to future shocks,

$$Var_t(y_{t+h}) = \phi_0\Sigma\phi'_0 + \phi_1\Sigma\phi'_1 + \cdots + \phi_{h-1}\Sigma\phi'_{h-1}$$

3.3 Forecasting accuracy measurements

The predicting performance of the ARMA and VAR model is compared by measuring forecasting accuracy statistics like Mean Forecasting Error (MFE) and Root Mean Square Forecast Error (RMSFE). The lower the calculated value of MFE, RMSFE, the better the model and hence, the forecasted values are considered accurate. Mathematically,

The mean forecast error (MFE) for horizon h is given below:

$$MFE_h = \frac{1}{T_h} \sum_{t=T_1+1}^{T-h} (y_{t+h} - y_{t,h}^f)$$

The root mean squared forecast error (RMSFE) for horizon h is:

$$RMSFE_h = \sqrt{\frac{1}{T_h} \sum_{t=T_1+1}^{T-h} (y_{t+h} - y_{t,h}^f)^2}$$

Where, $T_h = T - T_1 - h + 1$

Moreover, I applied the Diebold-Mariano test for equal forecast accuracy.

Forecast errors from two competing models (ARMA & VAR) are:

$e_{1t,h} = y_{t+h} - y_{1t,h}^f$ and $e_{2t,h} = y_{t+h} - y_{2t,h}^f$

The quadratic loss differential:

$$d_{t,h} = {e_{1t,h}}^2 - {e_{2t,h}}^2$$

The null of equal forecast accuracy:

H0: $\mathrm{E}(d_{t,h}) = 0$.

Test statistic of Diebold-Mariano (DM) is:

DM = $\frac{\overline{d_{t,h}}}{\sqrt{\frac{S}{T_h}}} \sim N(0,1)$

4. Results and discussions for the ARMA model

4.1 Time series plot

Data preprocessing is important for forecasting time series datasets. We have extracted the higher inflation period from my database to model and decide whether it plays an important role in shaping inflation. We use the vintage database for the study of forecasting inflation. After filtering and clearing my dataset, We figure out that the use of preprocessing data reduces the risk of computational problems as well as makes the training procedure more efficient. Figure 1 indicates the time series plot of the Polish HICP series from January 2002 to December 2020. The series exhibits a fluctuating pattern over time. It is seen that our time series data looks non-stationary. The mean does not seem to be constant over time and the series displays an upward and downward trend in time. In the below figure, the x-axis is the year, and the y-axis represents the Polish HICP inflation rate. Source: Figure 1 was created using R language (version 3.6.3) in R Studio, and the same process was applied to all other figures in this chapter.

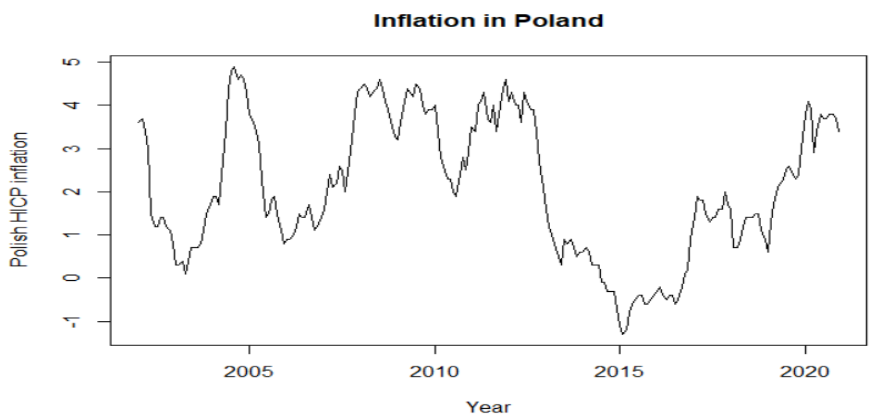

Figure 1: Time series plot of Polish HICP

4.2 Auto Correlation Function (ACF)

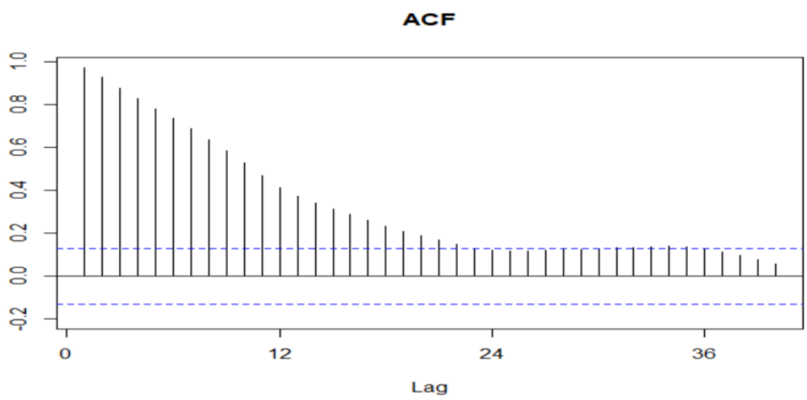

Figure 2: ACF plot of the inflation data

From Figure-2, We can see that the spikes of the ACF cross the cut-off line, suggesting that the current level of HICP is significantly autocorrelated with its lagged values. So, there is autocorrelation in the series. For our data series, it is evident that most of the lags fall outside the 95% confidence interval level, and lags from the first one up to the 23rd are statistically significant.

4.3 Partial Auto Correlation Function (PACF)

In Figure-3, we observe that the 1st lag and 2nd lag are statistically significant in the data series, while all other lags either lack statistical significance or show only marginal significance, notably at the 13th and 25th lags. This pattern is visually represented in the PACF plot, which exhibits two prominent spikes at lag 1 and lag 2, effectively indicating the points where the data deviates significantly from the baseline.

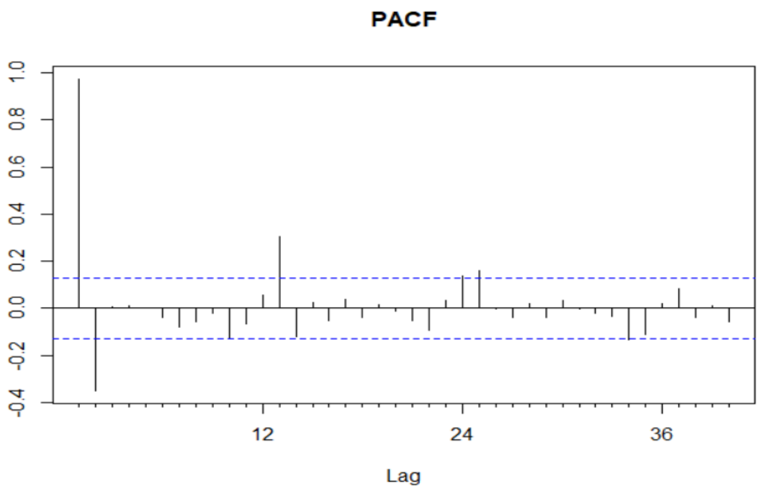

Figure 3: PACF plot of HICP series

4.4 Unit Root (UR) test

Here, we will apply the statistical test called Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test to see whether our data series is stationary or not.

By applying the Augmented Dickey-Fuller (ADF) test, we can see the following results. Source: the given table has been made by using Microsoft Office carefully and we follow the same manner to create other tables as well.

Table 1: ADF test of original series

Level	ADF statistic	Critical values (5% sig.)	Decision
Drift (intercept)	-2.49	-2.88	Non-Stationary

From Table 1, it is seen that the value of test-statistic t is: -2.49 and DF critical value is -2.88 at a 5% significant level. Since the calculated statistic is higher than the critical value for the ADF test, the null (H0) about non-stationary cannot be rejected. Hence, the time series is non-stationary in the optics of the ADF test.

By applying Phillips-Perron (PP) test, we see the following results of our data.

Table 2: PP test of original series

Level	PP statistic	Critical values (5% sig.)	Decision
Constant (intercept)	-2.20	-2.87	Non-Stationary

From Table 2, it can be said that the value of test-statistic t is: -2.20 and PP critical value is -2.87 at a 5% significant level. Since the calculated statistic is higher than the critical value for the PP test, the null (H0) about non-stationary cannot be rejected. Hence, the time series is non-stationary in the optics of the PP test.

After conducting unit root (UR) testing, we observed that utilizing non-stationary time series data is more suitable for modeling our restricted dataset. Using stationary time series data poses challenges in our cases due to the lower power of UR tests, indicating high persistence in annual inflation and forecasts stabilizing at a high value without returning to the mean.

4.5 ARMA model specification

We start by estimating ARMA (2, 2) model.

Table 3: Three methods result using ARMA (2, 2)

Methods	ar1	ar2	ma1	ma2	Constant
CSS-ML	0.584	0.353	0.7617	0.263	2.285
CSS	0.257	0.665	1.088	0.353	1.926
ML	0.200	0.729	1.102	0.303	2.316

Referring to Table-03, we prioritize simplicity in model selection, aiming for a minimal number of parameters. We prefer straightforward models; for instance, if a one-lag model adequately describes inflation, we favor it over a two-lag model, even if the second parameter is close to zero or statistically insignificant. My objective is to limit the complexity of the models while ensuring a good fit to the data.

However, it's important to note that different methods can yield varying results. Despite this variability, the parameters we obtain are reliable. We initiated the modeling process with the full-maximum likelihood approach, which provides a solid starting point. Subsequently, we employ conditional maximum likelihood (OLS estimator) and then return to the full maximum likelihood estimation (ML).

In these models, a constant term holds significance. This constant represents an equilibrium value, which plays a pivotal role in forecasting and ARMA modeling. Specifically, in this model, the equilibrium value stands at 2.285 (for instance), indicating that inflation in this context gradually returns towards this equilibrium. It's worth noting that the equilibrium values differ across various cases, signifying distinct patterns of equilibrium return.

4.6 Information criteria

Table 4: Bayesian information criteria (BIC)

Coefficient	ma0	ma1	ma2	ma3
ar0	860.32	595.96	425.19	333.99
ar1	183.57	158.62	160.21	165.61
ar2	154.85	160.27	165.54	170.71
ar3	160.07	163.50	170.07	166.24

Table5: Akaike information criteria (AIC)

Coefficient	ma0	ma1	ma2	ma3
ar0	853.46	585.67	411.47	316.84
ar1	173.30	144.92	143.08	145.06
ar2	141.17	143.17	145.02	146.77
ar3	142.99	143.01	146.16	138.91

Table 6: Hannan-Quinn information criteria (HQIC)

Coefficient	ma0	ma1	ma2	ma3
ar0	853.10	586.59	412.68	318.15
ar1	176.29	150.83	150.22	153.58
ar2	146.78	150.11	152.38	154.93
ar3	150.06	152.23	155.19	157.07

Referring to Table-4(a-c), we determine the appropriate lag structure by selecting a maximum of three lags. We use three different information criteria—Akaike Information Criteria (AIC), Normalized Bayesian Information Criteria (BIC), and Hannan-Quinn Information Criteria (HQIC)—to identify the best model. Based on the information criteria, the results are as follows: BIC suggests that ARMA (2, 0) is the best model with the lowest value of 154.85, AIC favors ARMA (3, 3) with a value of 138.91, and HQIC points to ARMA (2, 0) with a value of 146.78.

In summary, ARMA (3, 3) is a larger model with three lags, while ARMA (2, 0) is a more parsimonious model with only two lags. Ultimately, ARMA (3, 3) emerges as a good choice for analyzing and predicting inflation in this context.

4.7 Likelihood ratio (LR) test

After performing the LR test, we obtained a test statistic of 10.093 with a p-value of 0.038, which is less than the significance level of 0.05. This indicates a significant difference between the larger ARMA model and the smaller ARMA model.

The restriction imposed on the model is indeed having a significant impact on the fit to the data. In other words, the larger ARMA model and the smaller ARMA model show notably different fits to the data as a result of this restriction.

4.8 Estimating ARMA (2, 0) & ARMA (3, 3)

We am now comparing the ARMA (2, 0) and ARMA (3, 3) models to determine which one is a better fit.

Opting for the smaller ARMA (2, 0) model offers simplicity with only two parameters, making them easier to interpret. Conversely, the larger ARMA (3, 3) model introduces more complexity, with parameters like the first auto regression parameter (ar1) taking values such as -0.938 and 0.865, which can be challenging to intuitively understand. Additionally, working with a larger model can be more sophisticated due to the intricacies of its moving average representation.

Table 7: Results from two ARMA models

Model	Log-likelihood	AIC	AICc	BIC	ME	RMSE	MAE	MPE	MAPE	MASE	AF1
ARMA(2,0,0)	-66.5	141	141	155	-0.005	0.322	0.242	0.104	21.4	0.176	0.005
ARMA(3,0,3)	-61.5	139	140	166	-0.006	0.314	0.238	0.092	20.1	0.173	0.020

In Table 5, we present the results obtained from ARMA models with varying orders. To identify the best model, we consider several metrics, including log-likelihood, AIC, AICc, BIC, ME, RMSE, MAE, MPE, MAPE, MASE, and AF1. In general, a superior model exhibits the highest log-likelihood and the lowest values for AIC, AICc, BIC, ME, RMSE, MAE, MPE, MAPE, MASE, and AF1. Our aim is to minimize all these metrics when forecasting HICP inflation in Poland.After careful evaluation, the ARMA (2, 0) model emerges as the optimal choice. It not only meets the aforementioned criteria but also adheres to the parsimony principle, favoring the inclusion of the fewest parameters in the model. Therefore, we confidently select the ARMA (2, 0) model for our forecasting purposes.

4.9 Diagnostic checking of residuals

Figure-04(a, b) provides a visual diagnostic assessment of the ARMA (2, 0) model's fit to the time series data. Notably, the residuals of this model exhibit clear modeling behavior toward the end of the sample, indicating that standard errors remain consistent in both mean and variance over time.

Additionally, the ACF and PACF plots of the residuals demonstrate white noise characteristics, with exceptions being significant spikes at lag 12 for ACF and at lag 12 and lag 24 for PACF. These spikes could be attributed to random factors, possibly related to base effects. However, the Q-Q plot reveals that the residuals are not normally distributed, displaying a lack of fit in the tails of the distribution.

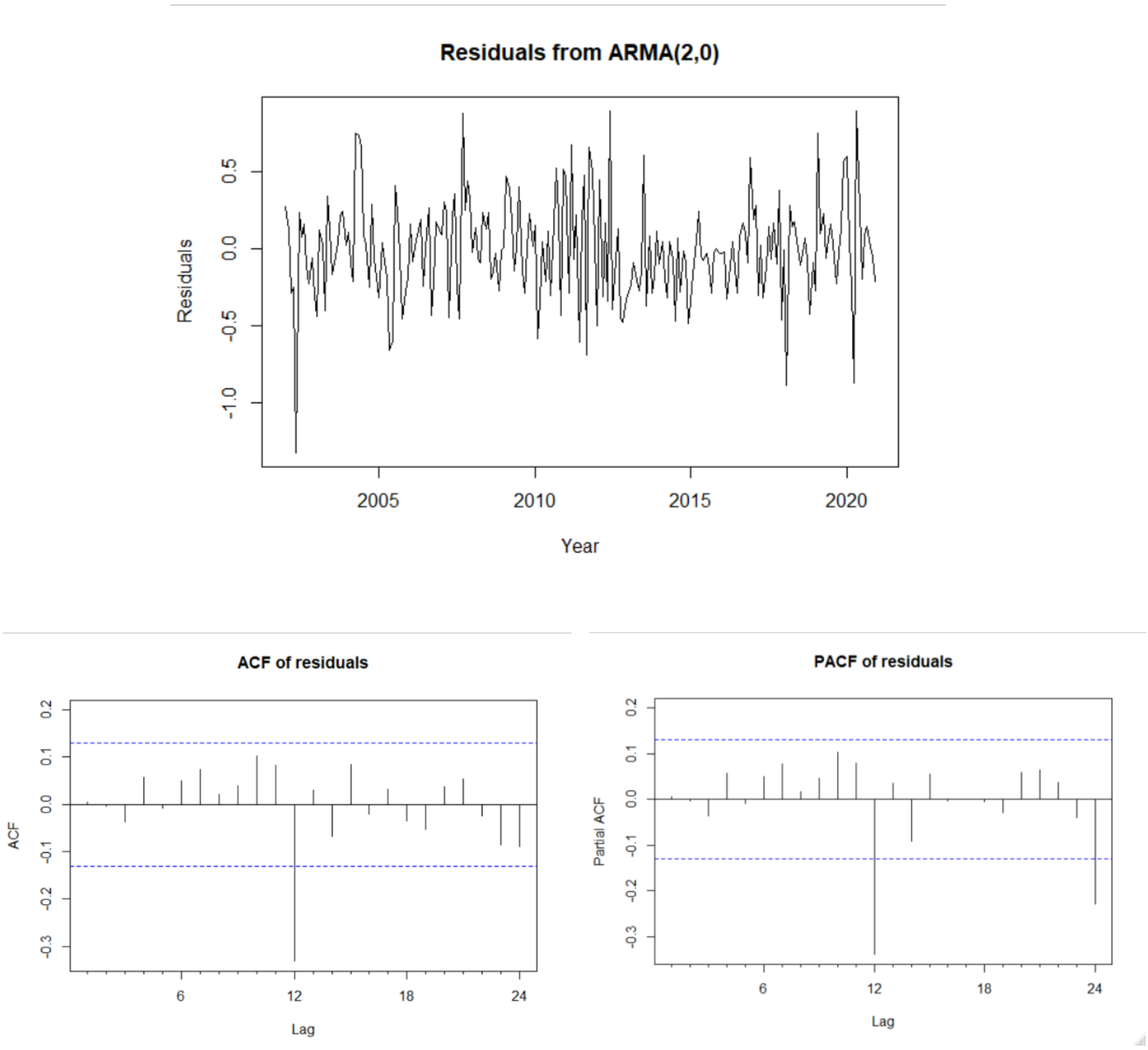

Figure 4: Correlograms plots of ARMA (2, 0) model

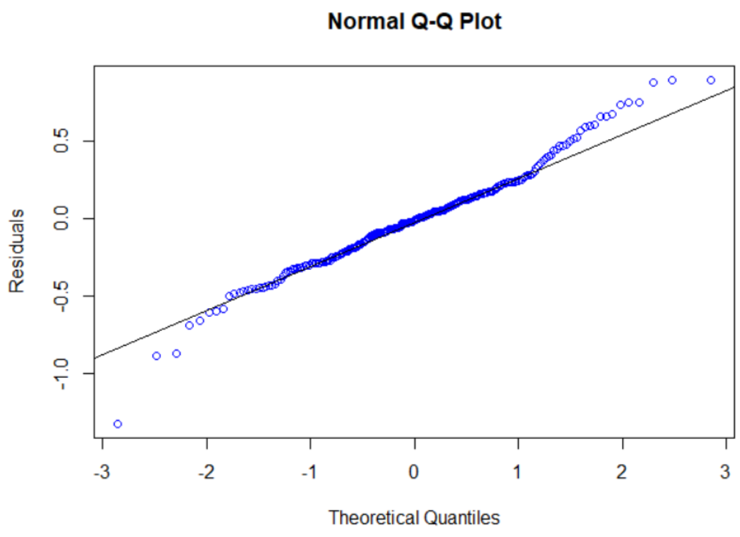

Figure 5: Normal Q-Q plot of residuals-ARMA (2, 0)

We conducted an Autocorrelation (Ljung-Box) test to assess the overall adequacy of the ARMA (2, 0) model. The test examined whether there was any serial autocorrelation in the model residuals. The results, with p-values of 0.6 for horizon h = 3, indicate that there is no statistically significant autocorrelation in the residuals.

This assessment was extended up to 12 lags, revealing that most lags did not exhibit serial autocorrelation, except for h = 12 (as shown in Table-6).

Table 8: Ljung-Box test table

Horizon	LB Stat.	p-value
h=3	0.311	0.577
h=4	1.065	0.587
h=5	1.082	0.781
h=6	1.687	0.793
h=7	2.994	0.701
h=8	3.107	0.795
h=9	3.490	0.836
h=10	6.007	0.646
h=10	7.685	0.566
h=12	34.306	0.000

Furthermore, we apply the two normality tests of the residuals for ARMA (2, 0). By Shapiro-Wilk and Anderson-Darling normality test, we can see that the residuals are not normally distributed (table-7)

Table 9: Test of normality-ARMA (2, 0)

Test of normality	Statistics	p-value
Shapiro-wilk	0.98028	0.002857
Anderson-Darling	1.0731	0.007976

In figure 5, the histogram of the residuals displays the constant mean and variance over time graphically.

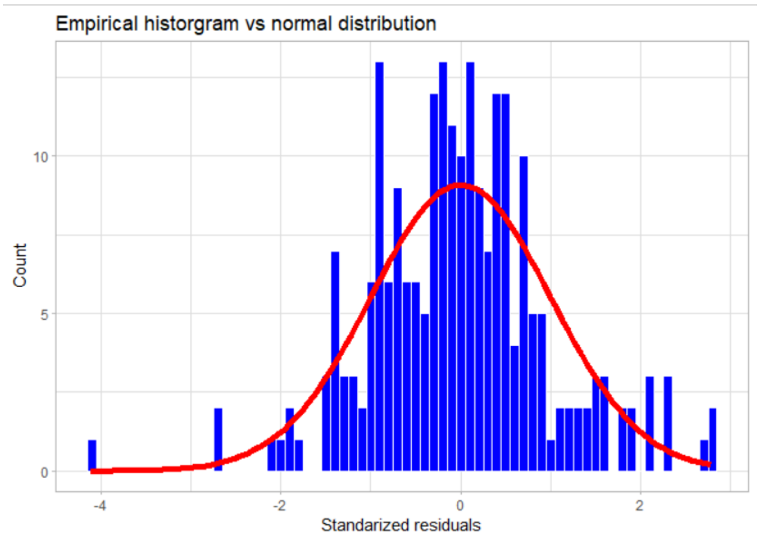

Figure 6: Histogram of residuals-ARMA (2, 0)

4.10 Out-of-sample: Forecasting inflation for the next 5 years from ARMA (2, 0)

Using the ARMA (2, 0) model, we have illustrated how inflation forecasts revert to equilibrium more rapidly, as depicted in Figure 6. Over the observed time periods, inflation displays fluctuations. In 2021, the predicted inflation stands at 2.952, but it gradually decreases continuously until 2025.

For instance, in 2023, it is expected to be 2.421, and in 2024, it is projected to reach 2.334. In the graph, the blue line represents predicted inflation, while the green line represents the mean value of the dataset. The x-axis corresponds to the year, while the y-axis denotes the inflation rate (HICP). Between 2021 and 2025, a notable trend emerges: the predicted inflation converges quickly toward the mean value.

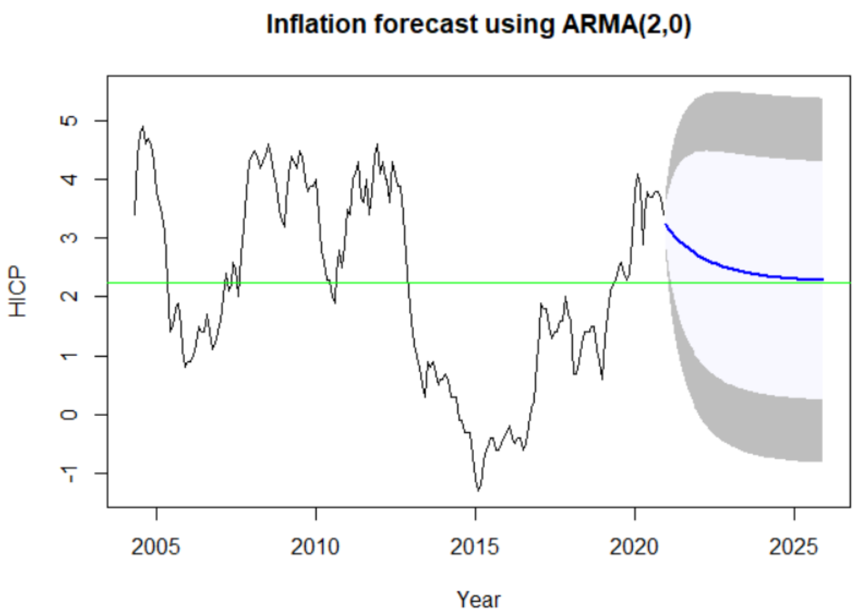

Figure 7: Forecasting graph using ARMA (2, 0)

From table-08, we can see that in Poland inflation will be lower and it will return to equilibrium quickly after 2024 years. In the table below, the bold font denotes forecasts dates and rates.

Table 8: Forecasting results-ARMA (2, 0)

Year	HICP (x)	Year	HICP (x)
2002-12-01	1.958	2014-12-01	0.083
2003-12-01	0.725	2015-12-01	-0.700
2004-12-01	3.633	2016-12-01	-0.192
2005-12-01	2.183	2017-12-01	1.617
2006-12-01	1.258	2018-12-01	1.192
2007-12-01	2.617	2019-12-01	2.117
2008-12-01	4.167	2020-12-01	3.667
2009-12-01	4.025	**2021-12-01**	**2.952**
2010-12-01	2.658	**2022-12-01**	**2.595**
2011-12-01	3.900	**2023-12-01**	**2.421**
2012-12-01	3.700	**2024-12-01**	**2.334**
2013-12-01	0.817	**2025-12-01**	**2.291**

In conclusion, our analysis demonstrates that ARMA (2, 0) provides a superior model for inflation prediction. Notably, it yields forecasts that exhibit a steady return to the mean, contributing to economic stability. Additionally, our model performs satisfactorily during periods of low inflation. Furthermore, our analysis underscores the robustness of our findings across various model specifications considered. These results collectively support the effectiveness and reliability of the ARMA (2, 0) model for inflation forecasting.

5. Results and discussions for the VAR model

5.1 Time series plot between PL and EA

Figure 7 illustrates the fluctuating pattern of inflation over time for both Poland (PL) and the Euro Area (EA). Both series exhibit alternating upward and downward trends. The x-axis represents the year, while the y-axis depicts the Polish HICP inflation rate. Source: The figure was generated using R language (version 3.6.3) in R Studio, following the same procedure used for all other figures in this chapter.

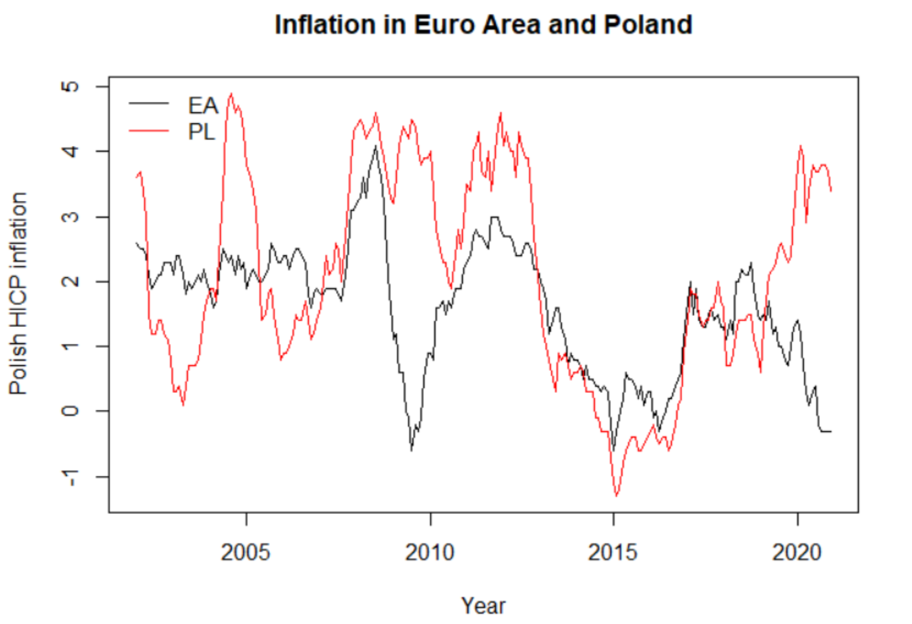

Figure 8: Time series plot of HICP

5.2 Information criteria

From table-9 of the information criteria, we can say that 2 lag is sufficient for a VAR model.

Table 9: information criteria

Selection			
AIC(n)	HQ(n)	SC(n)	FPE(n)
5	5	2	5

Criteria						
	1	2	3	4	5	6
AIC(n)	-4.939	-5.080	-5.120	-5.150	-5.233	-5.219
HQ(n)	-4.902	-5.019	-5.033	-5.039	-5.097	-5.058
SC(n)	-4.847	-4.927	-4.905	-4.874	-4.896	-4.820
FPE(n)	0.007	0.006	0.006	0.005	0.005	0.005

5.3 Estimating parameters for the VAR model

In Table-10 (a), the equation for Euro Area inflation reveals that one estimated parameter stands out as statistically significant, indicated by a star mark (*) due to p-values approaching zero, which allows us to reject the null hypothesis at a 5% significance level. Specifically, the autoregression coefficients are 1.090 (statistically significant) and -0.120. Additionally, there are coefficients linked to inflation in Poland (PL), with 0.023 indicating a slight positive impact and -0.030 suggesting a minor reduction. Notably, the coefficient for inflation in PL is much lower than the autoregression coefficients. The high multiple R-squared value of 0.932 indicates a strong correlation between the response and the fitted values. Source: The table was meticulously created using Microsoft Office, following the same methodology applied to other tables in this study.

Estimation results for equation EA:

EA = EA. l1 + Pl. l1 + EA. l2 + PL. l2 + Constant

Table 10: the equation of inflation for the Euro Area

Variables	Estimate Std.	Error	t value	Pr(>|t|)
EA. l1	1.090	0.070	15.44	$<2 \times 10^{-16}$
Pl. l1	0.023	0.052	0.45	0.652
EA. l2	-0.120	0.071	-1.68	0.094
PL. l2	-0.030	0.052	-0.57	0.570
Constant	0.050	0.035	1.43	0.153

In Table 10(b), the equation for inflation in Poland reveals that all estimated parameters are statistically significant, with p-values close to zero, allowing us to reject the null hypothesis at both 1% and 5% significance levels. Specifically, the coefficients include 0.200 for the inflation rate in the Euro Area, which is higher than -0.2046. Additionally, we find statistically significant coefficients of 1.2948 and -0.3258. Each of these coefficients proves to be individually useful for prediction. The high adjusted R-squared value of 0.96 suggests that additional input variables (independent variables) significantly contribute to the VAR model, accounting for 96% of the variance.

Estimation results for equation PL:

PL = EA. l1 + Pl. l1 + EA. l2 + PL. l2 + Constant

Table 11: the equation of inflation for Poland

Variables	Estimate Std.	Error	t value	Pr(>|t|)
EA. l1	0.200	0.087	2.29	0.023
Pl. l1	1.294	0.065	19.83	$<2 \times 10^{-16}$
EA. l2	-0.204	0.089	-2.30	0.022
PL. l2	-0.325	0.065	-4.96	1.4×10^{-6}
Constant	0.071	0.043	1.65	0.100

5.4 Model checking

Figures 8(a, b) display the graphical diagnosis of the VAR (2) model to assess its fit to the time series data. The residuals of this model reflect the sample modeling. In terms of the residuals' behavior, both the ACF plots for PL and EA exhibit white noise characteristics. However, notable spikes are observed at lag 12 for both PL and EA, likely attributed to base effects. On the other hand, the Q-Q plot reveals that the residuals do not follow a normal distribution, displaying a lack of fit in the distribution's tails.

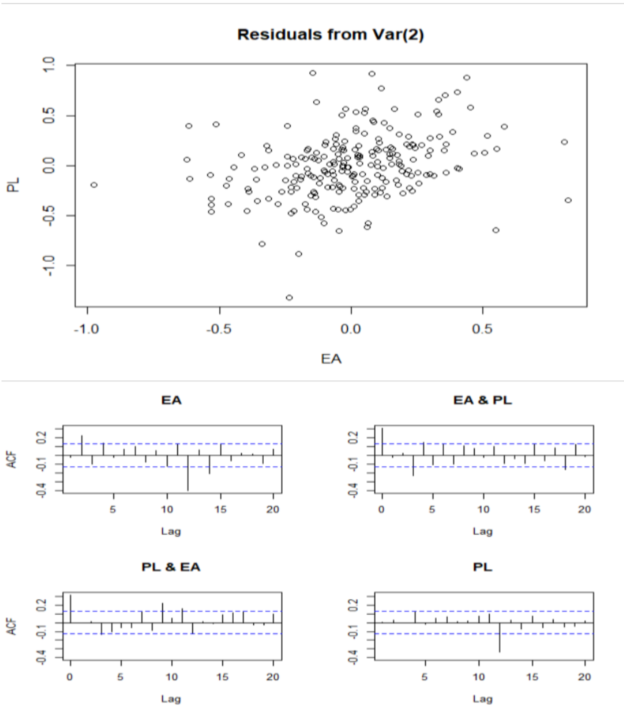

Figure 9: Diagnostic plot of residuals-VAR (2)

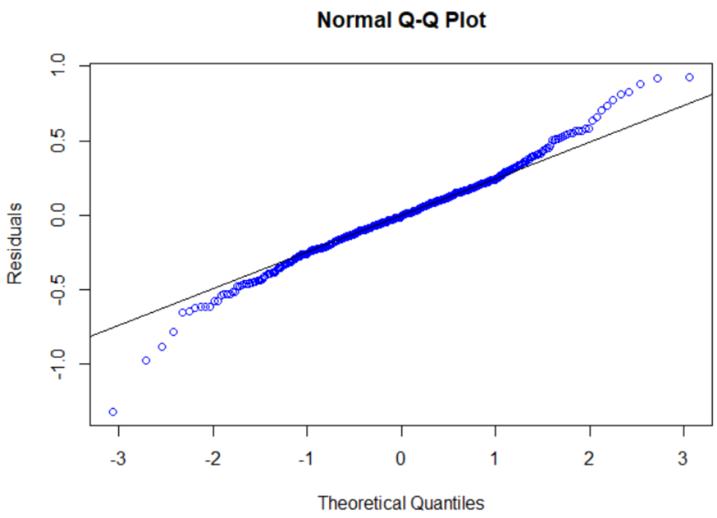

Figure 10: Normal Q-Q plot- VAR (2)

We conducted an Autocorrelation (Ljung-Box) test with a lag length of 6 to assess serial autocorrelation in the residuals of the VAR (2) model. The p-value (8.878×10^(-07)) is less than 0.05, indicating the presence of autocorrelation. Consequently, finding a VAR model with no autocorrelation is challenging, and options include models with 2 or 5 lags.Additionally, we applied two normality tests (Shapiro-Wilk and Anderson-Darling) to the residuals of the VAR (2) model, both of which indicate that the residuals are not normally distributed (as shown in Table-11).

Table 12: Test of normality-VAR (2)

Test of normality	Statistics	p-value
Shapiro-wilk	0.98504	0.0001316
Anderson-Darling	1.4449	0.0009781

Also, in figure 9, the histogram of the residuals shows the distribution of errors over time graphically.

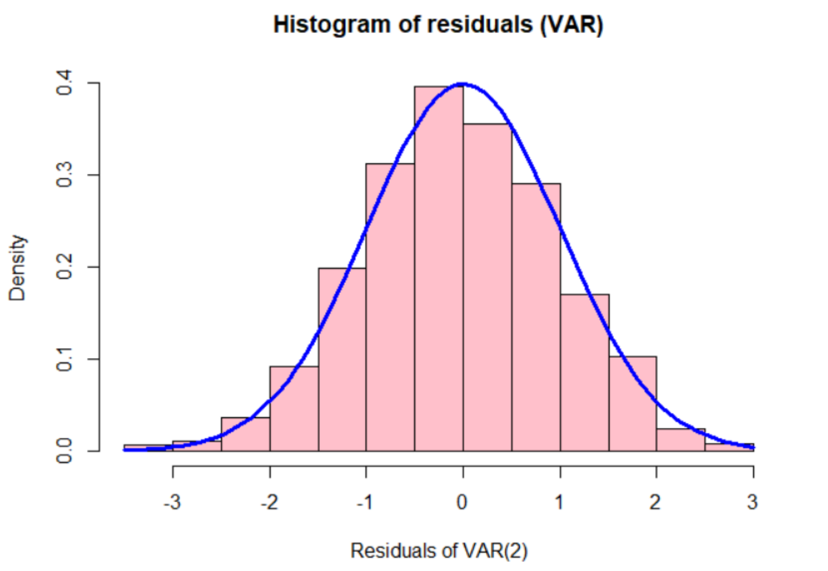

Figure 11: Histogram of residuals-VAR (2)

5.5 Out-of-sample: Forecasting inflation for Euro Area and Poland for the next 5 years using VAR (2) model

Figure-10 reveals a gradual trend in forecasting inflation for Poland. In 2021, the predicted inflation is 3.082, but it steadily decreases until 2025. Over the next four years, inflation is expected to remain low, with forecasts of 2.643 for 2023 and 2.506 for 2024. In the graph, the x-axis corresponds to the year, while the y-axis represents the inflation rate (HICP). The straight line represents the mean value, and the blue line indicates the inflation rate. Notably, between 2021 and 2025, the predicted inflation gradually converges toward the mean value.

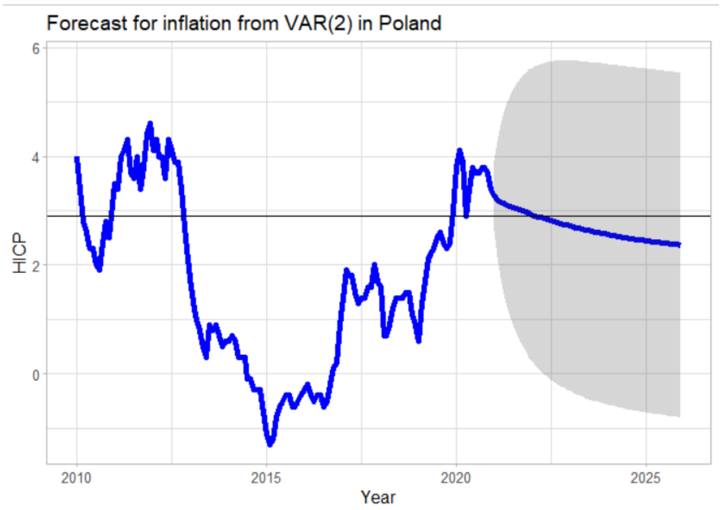

Figure12: Forecasting graph using VAR (2) for Poland

From table-12, we can see that in Poland inflation will be lower and it will return to equilibrium quickly after 2024 years. The bold font indicates forecasts in the given table.

Table 13: Forecasting results-VAR (2) for Poland

Year	HICP (x)	Year	HICP (x)
2002-12-01	1.958	2014-12-01	0.083
2003-12-01	0.725	2015-12-01	-0.700
2004-12-01	3.633	2016-12-01	-0.192
2005-12-01	2.183	2017-12-01	1.617
2006-12-01	1.258	2018-12-01	1.192
2007-12-01	2.617	2019-12-01	2.117
2008-12-01	4.167	2020-12-01	3.667
2009-12-01	4.025	**2021-12-01**	**3.082**
2010-12-01	2.658	**2022-12-01**	**2.827**
2011-12-01	3.900	**2023-12-01**	**2.643**
2012-12-01	3.700	**2024-12-01**	**2.506**
2013-12-01	0.817	**2025-12-01**	**2.407**

Figure-11 depicts a gradual trend in forecasting inflation for the Euro Area. In 2021, the predicted inflation is -0.066, but it steadily increases until 2025. Over the next four years, inflation is expected to remain very low, with forecasts of 0.571 for 2023 and 0.758 for 2024. In the graph, the x-axis corresponds to the year, while the y-axis represents the inflation rate (HICP). The straight line represents the mean value, and the blue line indicates the inflation rate. Notably, between 2021 and 2025, the predicted inflation slowly converges toward the mean value.

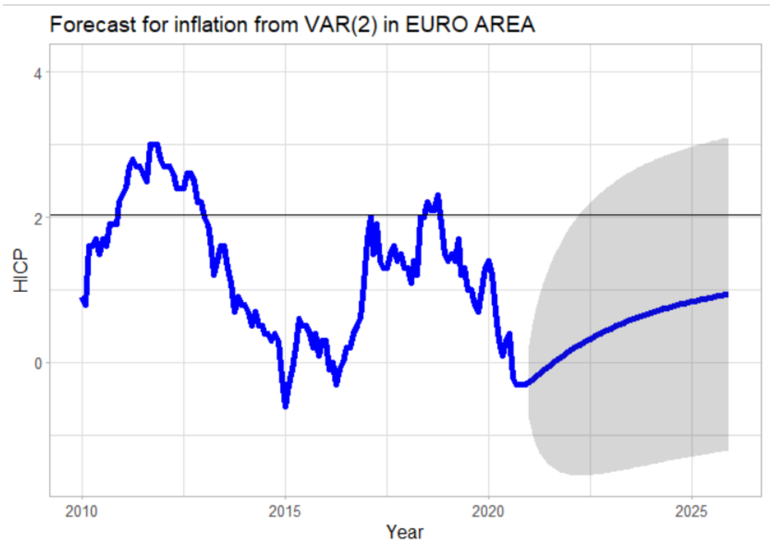

Figure 13: Forecasting graph using VAR (2) for Euro Area

From table-13, we see that in the euro area inflation will be lower and it will return to equilibrium slowly after 2024 years. The bold font indicates forecasts in the given table.

Table 14: Forecasting results-VAR (2) for Euro Area

Year	HICP (x)	Year	HICP (x)
2002-12-01	1.958	2014-12-01	0.083
2003-12-01	0.725	2015-12-01	-0.700
2004-12-01	3.633	2016-12-01	-0.192
2005-12-01	2.183	2017-12-01	1.617
2006-12-01	1.258	2018-12-01	1.192
2007-12-01	2.617	2019-12-01	2.117
2008-12-01	4.167	2020-12-01	3.667
2009-12-01	4.025	**2021-12-01**	**-0.066**
2010-12-01	2.658	**2022-12-01**	**0.307**
2011-12-01	3.900	**2023-12-01**	**0.571**
2012-12-01	3.700	**2024-12-01**	**0.758**
2013-12-01	0.817	**2025-12-01**	**0.890**

In summary, our analysis indicates that inflation prediction using VAR (2) yields favorable results. However, it's noteworthy that the forecast values gradually converge toward the mean value for both Poland (PL) and the Euro Area (EA), a contrast to the faster convergence observed with ARMA (2, 0). We have presented the predicted results comprehensively and accurately throughout our analysis. Ultimately, our findings highlight the robustness and effectiveness of the analysis across various model specifications.

6. The comparison table of ARMA (2, 0) and VAR (2)

Table-14 demonstrates that both ARMA and VAR models exhibit a trend of predicted inflation converging toward equilibrium when compared to the years 2021 and 2022. Notably, the ARMA model's forecast results show a faster return to the mean value compared to the VAR model.

Table15: Forecast comparison table of the AMRA (2, 0) and VAR (2) models

Year	Actual inflation	ARMA (2, 0)	VAR (2)
2002-12-01	0.8	1.958	1.958
2003-12-01	1.7	0.725	0.725
2004-12-01	4.3	3.633	3.633
2005-12-01	0.8	2.183	2.183
2006-12-01	1.4	1.258	1.258
2007-12-01	4.3	2.617	2.617
2008-12-01	3.3	4.167	4.167
2009-12-01	3.9	4.025	4.025
2010-12-01	2.9	2.658	2.658
2011-12-01	4.6	3.900	3.900
2012-12-01	2.1	3.700	3.700
2013-12-01	0.6	0.817	0.817
2014-12-01	-0.7	0.083	0.083
2015-12-01	-0.4	-0.700	-0.700
2016-12-01	0.9	-0.192	-0.192
2017-12-01	1.7	1.617	1.617
2018-12-01	0.9	1.192	1.192
2019-12-01	3.0	2.117	2.117
2020-12-01	3.4	3.667	3.667
2021-12-01	8.0	**2.952**	**3.082**
2022-12-01	15.3	**2.595**	**2.827**
2023-12-01	-	**2.421**	**2.643**
2024-12-01	-	**2.334**	**2.506**
2025-12-01	-	**2.291**	**2.407**

6.1 Comparison between the ARMA model and VAR model graphically

Figure-12 provides a graphical comparison of the proposed models with actual inflation, focusing on their predicted inflation rates. It is evident from the graph that the ARMA model is converging towards equilibrium at a faster pace than the VAR model. This comparison strongly suggests the superior performance of the ARMA (2, 0) model in the coming years. In the figure, the x-axis corresponds to the year in the database, while the y-axis represents the inflation rate. Source: The figure was created using R language (version 3.6.3) in R Studio.

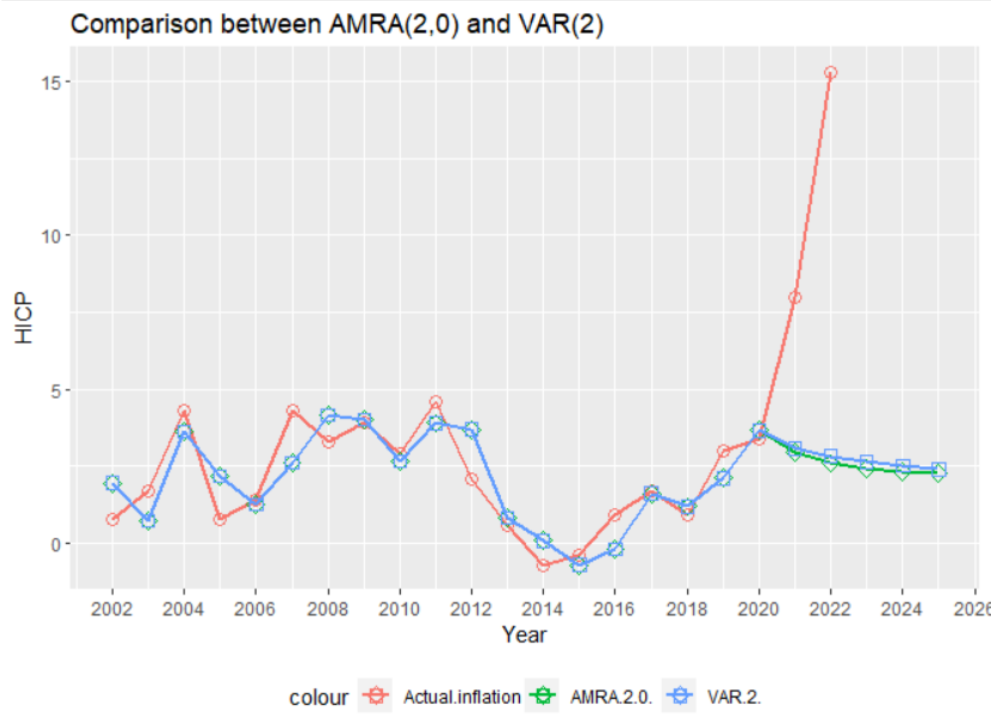

Figure14: Comparison graph

In summary, the ARMA (2, 0) model proposes lower inflation forecasts and exhibits a faster return to the mean value, which is a crucial aspect for accurate forecasting. In contrast, the VAR (2) model takes a slower approach to reach the mean value.

The graphical representation with three colors, where red represents actual inflation, green represents the ARMA (2, 0) model, and blue represents the VAR (2) model, vividly illustrates the differences in inflation predictions.

7. Evaluation to achieve the best model using various test

7.1 Mean forecast error (MFE) and Root mean squared forecast error (RMSFE)

We have initiated an analysis to assess the accuracy of forecasts using two commonly employed criteria: Mean Forecast Error (MFE) and Root Mean Squared Forecast Error (RMSFE). As seen in Table 15(a, b), when evaluated using the MFE criterion, the ARMA model consistently exhibits lower errors across most forecast horizons. Similarly, in terms of the Root Mean Squared Forecast Error, the ARMA model outperforms the VAR model, particularly at shorter horizons.

In summary, the comparative accuracy analysis reveals that the ARMA model provides satisfactory results for forecasting inflation, surpassing the VAR model in terms of accuracy. These findings underscore the advantages of employing the ARMA model over the VAR model.

Table 16: Mean forecast error (ME)

Horizon	1	2	3	6	9	12
ARMA	**0.0223**	**0.0585**	**0.0999**	**0.242**	**0.381**	**0.530**
VAR	0.0357	0.0844	0.1385	0.319	0.499	0.685

Table 17: Root mean squared forecast error (RMSE)

Horizon	1	2	3	6	9	12
ARMA	**0.314**	**0.519**	**0.682**	**0.849**	**0.889**	**1.07**
VAR	0.317	0.524	0.692	0.891	0.955	1.17

7.2 Diebold Mariano test

We conducted a Diebold-Mariano test with a horizon of 12, and the results allow us to reject the null hypothesis (H0) of equal forecast accuracy. The DM statistic (-4.45) falls outside the critical z-value range of -1.96 to 1.96, indicating that both the ARMA and VAR models do not exhibit equal forecast accuracy.

Furthermore, in Figure-13, we observe that the forecasting errors between the two models exhibit variations over time. The ACF plot suggests the presence of autocorrelation in errors at the beginning of the lag.

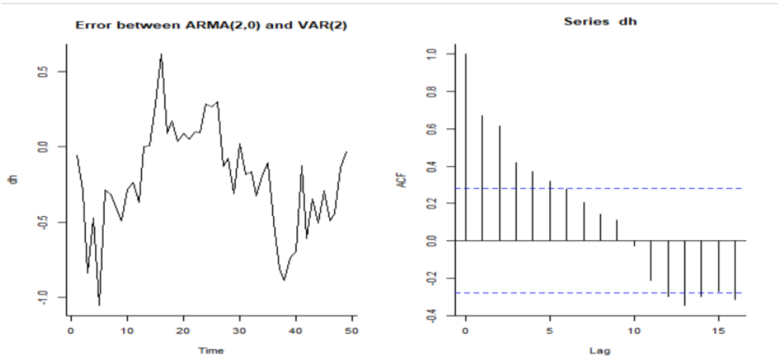

Figure 15: Errors difference

Considering the Diebold-Mariano (DM) test results, which indicate that both models do not have equal forecast accuracy, it is reasonable to infer that one model outperforms the other. The test results provide evidence that one of the models exhibits superior accuracy compared to the other.

7.3 Forecast accuracy measures: efficiency test for AMRA (2, 0)

We conducted an Efficiency/Unbiasedness test, involving regression between the actual value and forecast value. The results indicate that our forecast parameter (x level) is statistically significant, with a p-value of 0.001, which is less than 0.05 at a 5% significance level. Consequently, we can reject the null hypothesis (H0).

Furthermore, when subjected to a linear hypothesis test, the parameters remain statistically significant. These findings suggest that the unbiasedness criterion is not met.

7.4 Efficiency test - graphical illustration - AMRA (2, 0)

Figure-14 provides a clear graphical representation of the efficiency test for ARMA (2, 0). The scatterplot highlights a systematic underprediction trend, indicating that the model consistently underpredicts actual values.

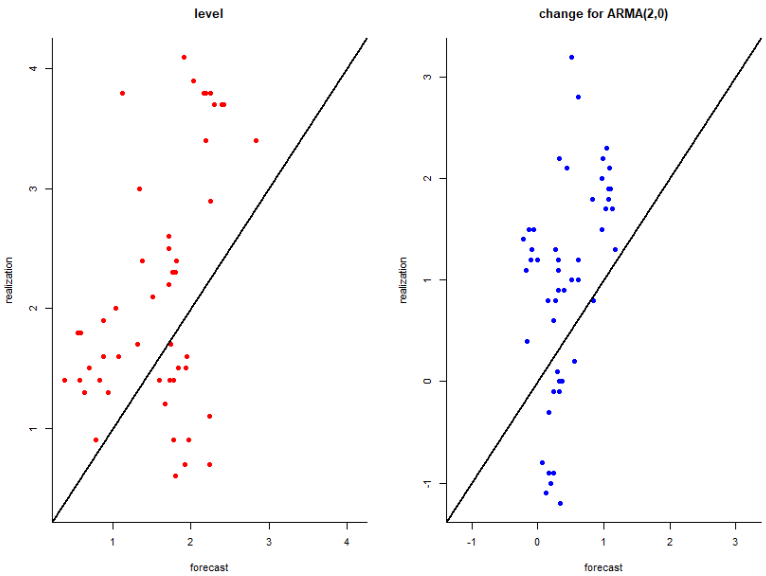

Figure 16: Scatter plot of efficiency test-ARMA (2, 0)

7.5 Forecast accuracy measures: efficiency test for VAR (2)

A relatively good forecast accuracy does not imply that they are satisfactory in the absolute sense. Absolute performance includes ME and efficiency tests. So, we manage the following efficient test.

From the Efficiency/unbiasedness test (regression between actual value and forecast value), we can say that our forecast parameter (x level) is statistically significant due to the p-value (0.002) being less than 0.05 at a 5% significant level. So, we can reject the null (H0). By the linear hypothesis test, once again parameters are statistically significant. Hence, we can see that unbiasedness is no met.

7.6 Efficiency test - graphical illustration – VAR (2)

In Figure-15, we provide a clear graphical representation of the efficiency test for the VAR (2) model. The scatterplot illustrates a consistent pattern of systematic underprediction, indicating that the VAR (2) model consistently underpredicts actual values.

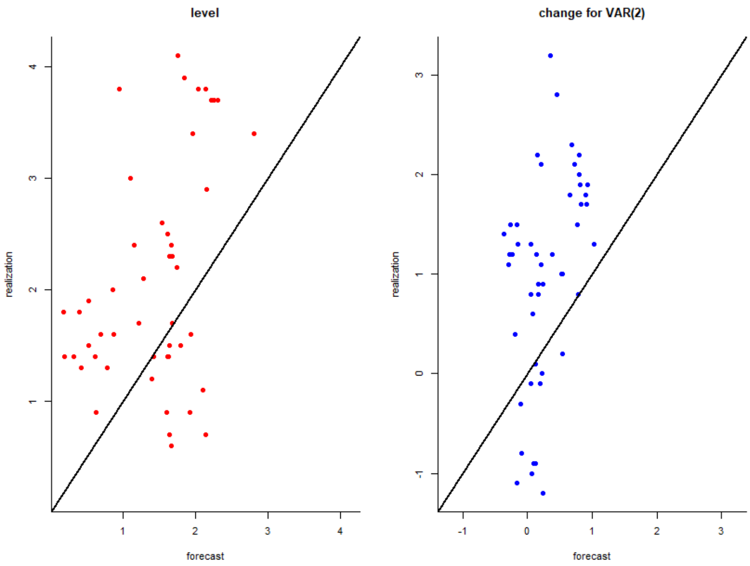

Figure 17: Scatter plot of efficiency test-VAR (2)

8. Sequential forecasts for ARMA (2, 0) & VAR (2)

In Figure-16(a), we present a visual representation of the rolling window forecast analysis using the ARMA (2, 0) model with a restricted dataset spanning from 2010 to 2024. The analysis highlights the model's ability to rapidly revert towards the sample mean when forecasting future events.

The x-axis corresponds to the year, while the y-axis represents the Polish inflation rate. It becomes evident that when inflation is high, it tends to return to equilibrium. Notably, when inflation is at 2, the model exhibits a faster convergence towards equilibrium compared to other scenarios. Conversely, when inflation is at 4, the reversion to equilibrium occurs more gradually.

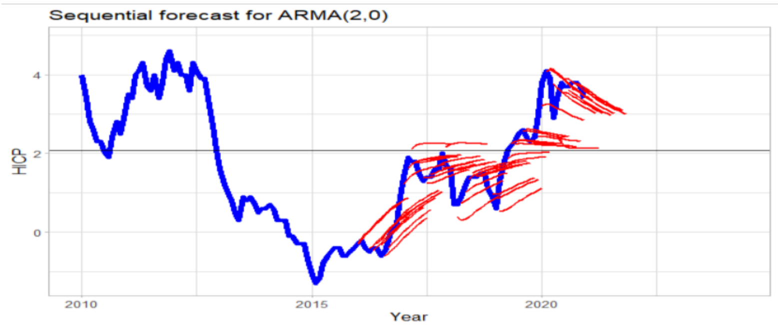

Figure 18: Sequential forecast-ARMA (2, 0)

In Figure-16(b), we provide a graphical representation of the rolling window forecast analysis using the VAR (2) model. While the VAR (2) model also exhibits reversion towards equilibrium, the dynamics can be more complex compared to the ARMA (2, 0) model. However, it is evident from the figure that there is still a discernible reversion towards the mean (equilibrium).

The graph illustrates that the VAR (2) model gradually moves towards the mean value over time.

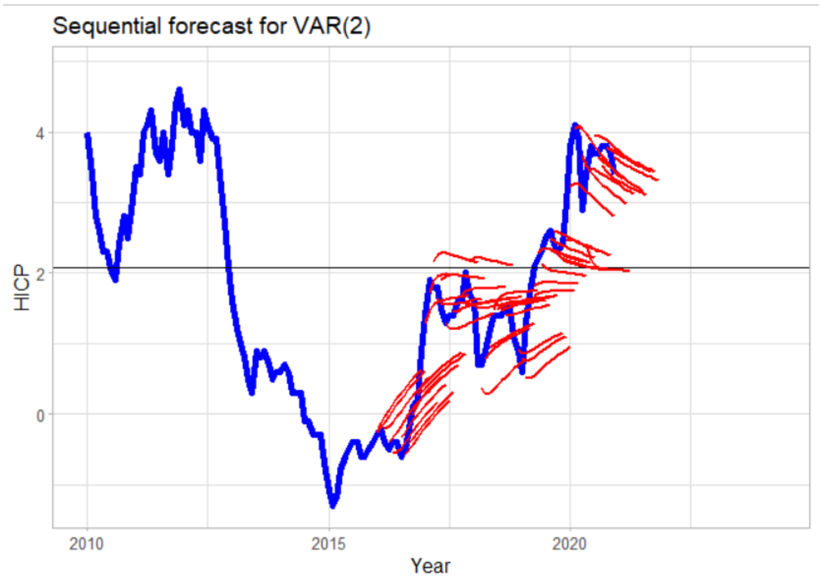

Figure 19: Sequential forecast-VAR (2)

9. Conclusion

This research aimed to model and forecast Polish inflation using two methodologies: the Autoregressive Moving Average (ARMA) model and the Vector Autoregressive (VAR) model. The study utilized annual data from January 2002 to December 2020 and conducted various diagnostic checks, graphical investigations, and statistical tests to assess the adequacy of the models. The comparison between the ARMA and VAR models provided valuable insights into their respective abilities to accurately predict the inflation rate. The evaluation, based on Mean Forecast Error (MFE) and Root Mean Squared Forecast Error (RMSFE), revealed that the ARMA (2, 0) model outperformed other models in forecasting inflation for the next five years. Additionally, the analysis indicated a downward trend in the Harmonized Index of Consumer Prices (HICP) for Poland over the forecasted period. As a result, this study suggests that fiscal and monetary authorities should consider implementing robust economic policies to maintain the inflation rate at an equilibrium level in the coming years.

9.1 Recommendations for future works

In future research, we intend to explore the application of Bayesian inference alongside the methods employed in this thesis for enhanced analysis and forecasting. Additionally, we plan to expand our repertoire by delving deeper into various methods and staying updated with the latest developments in the field.

The solution methodology presented in this thesis holds promise for addressing a wide range of real-life challenges in the finance and economic sectors. We believe that further research in this direction will yield valuable insights and contribute to the advancement of predictive modeling and analysis.

References

[1] Khan, S., & Alghulaiakh, H. (2020). ARIMA model for accurate time series stocks forecasting. *International Journal of Advanced Computer Science and Applications*, *11*(7).

[2] Szafranek K. Bagged neural networks for forecasting Polish (low) inflation. International *Journal of Forecasting*. 2019 Jul 1;*35*(3):1042-59.

[3] Hernandez-Matamoros, A., Fujita, H., Hayashi, T., & Perez-Meana, H. (2020). Forecasting of COVID-19 per region using ARIMA models and polynomial functions. *Applied soft computing*, *96*, 106610.

[4] Abonazel, M. R., & Abd-Elftah, A. I. (2019). Forecasting Egyptian GDP using ARIMA models. *Reports on Economics and Finance*, *5*(1), 35-47.

[5] Khan, Firdos, Alia Saeed, and Shaukat Ali. "Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan." *Chaos, Solitons & Fractals 140* (2020): 110189.

[6] Ramyar, S., & Kianfar, F. (2019). Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, *53*, 743-761.

[7] Gottschalk, J. and Moore, D. (2001) ‘‘Implementing inflation targeting regimes: the case of Poland,’’ *Journal of comparative economics*, Vol. *29*(1)

[8] Moser, G., Rumler, F. and Scharler, J. (2004.) "Forecasting Austrian Inflation," Working Papers 91, Oesterreichische National bank (Austrian Central Bank).

[9] Rumler, F., Moser, G. and Scharler, J. (2002), ‘‘Forecasting Austrian HCPI and its components using VAR and ARIMA Models’’ Oesterreichische National Bank, working paper 73.

[10] Uko, A. K and Nkoro, E (2012) ‘‘Inflation Forecasts with ARIMA, Vector Autoregressive and Error Correction Models in Nigeria,*’’European Journal of Economics, Finance and Administrative Sciences*, issue *50*.

[11] Poulos, Laurette, Alan Kvanli, and Robert Pavur. "A comparison of the accuracy of the Box-Jenkins method with that of automated forecasting methods." *International Journal of Forecasting 3*, no. *2* (1987): 261-267.

[12] Arratibel, O., Kamps, C., & Leiner-Killinger, N. (2009). Inflation forecasting in the new EU member states.

[13] Hubrich K. Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy? *International Journal of Forecasting*. 2005 Jan 1;*21*(1):119-36.

[14] Pufnik A, Kunovac D. Short-term forecasting of inflation in Croatia with seasonal ARIMA processes. 2006 Dec.

[15] Junttila J. Structural breaks, ARIMA model and Finnish inflation forecasts. *International Journal of Forecasting*. 2001 Apr 1;*17*(2):203-30.

[16] Wigati, Y., Rais, R., & Utami, I. T. (2015). Time Series Modeling with the ARIMA Process for Prediction of Consumer Price Index (CPI) In Palu – Central Sulawesi (Pemodelan Time Series Dengan Proses Arima Untuk Prediksi Indeks Harga Konsumen (IHK) Di Palu – Sulawesi Tengah). *Jurnal Ilmiah Matematika dan Terapan*, *12*(2), 149-159.

[17] Mohamed J. Time series modelling and forecasting of Somaliland consumer price index: a comparison

of ARIMA and regression with ARIMA errors. *American Journal of Theoretical and Applied Statistics*. 2020;*9*(4):143-53.

[18] Tandon H, Ranjan P, Chakraborty T, Suhag V. Coronavirus (COVID-19): ARIMA-based Time-series Analysis to Forecast Near Future and the Effect of School Reopening in India. *Journal of Health Management*. 2022 Sep;*24*(3):373-88.

[19] Djawoto, D. (2010). Advanced Forecasting of Inflation with Auto Regressive Integrated Moving Average (ARIMA) Method (Peramalan Laju Inflasi Dengan Metode Auto Regressive Integrated Moving Average (ARIMA)). *Jurnal Ekonomi dan Keuangan*, *14*(4), 524-538.

[20] Kumar M, Anand M. An application of time series ARIMA forecasting model for predicting sugarcane production in India. *Studies in Business and Economics*. 2014 Apr 1;*9*(1):81-94.