Theoretical and Applied Economics
Volume XXX (2023), No. 1(634), Spring, pp. 67-76

Forecasting the Romanian inflation rate: An Autoregressive Integrated Moving-Average (ARIMA) approach

Rareș-Petru MIHALACHE
National Institute for Economic Research "Costin C. Kirițescu", Romanian Academy, Romania
rares.mihalache9@gmail.com
Dumitru Alexandru BODISLAV
Bucharest University of Economic Studies, Romania
alex.bodislav@ase.ro

Abstract. *The primary objectives of this paper are to empirically create an univariate Autoregressive Integrated Moving-Average (model) using Box-Jenkins methodology to forecast Romanian inflation and inspect the prediction performance of the estimated model between October 2021 and October 2022. This study uses Ordinary Least Squares (OLS) technique for estimation purposes. On the foundation of different selection assessment and diagnostic criteria, the best model is selected to predict inflation in Romania in the short-run. We find that ARIMA (7, 1, 1) model is a suitable one under model identification, parameters estimation, diagnostic checking, and inflation prediction. In-sample forecasting is performed and the estimated ARIMA model reasonably tracks the actual inflation in the sample period.*

Keywords: ARIMA, Box-Jenkins, inflation forecasting, time-series modelling, National Bank of Romania (NBR).

JEL Classification: C010, C51, C520, C530, E3.

1. Introduction

Inflation represents one of the most significant macroeconomic variables and the most worried by the economic agents, including the government, because it brings severe influence on the level of the welfare and on the architecture of the production costs. Some of the nations which experienced hyperinflation highlighted that poor inflation would yield to political and social instability (Yolanda, 2017).

Inflation is defined as an important macroeconomic imbalance manifested by a general growth in the price level over an interval of time, resulting in a constant decrease in the purchasing power of money. When there is a general rise in prices, each currency unit purchases fewer services and goods. As a result, inflation displays a decline in the purchasing power of the money (Csorba and Juravle, 2022). There are two main categories of inflation: demand pull inflation and cost push inflation. Demand pull inflation defines inflation where the root cause comes from the demand part. The constant growth in demand is determined by factors such as an increase in money supply, a growth in government purchase, increases in exports, etc. When demand is increased and it cannot be satisfied by a similar increase in supply, the general price level will raise and inflation will happen. On the other hand, cost push inflation, also called supply push inflation, happens due to a rise in the production cost. For instance, an increase of the raw materials price, a growth of wage rate, etc. The general price level of services and goods will increase when there is a rise of production costs within industries (Majumder, 2016).

Monetary authorities have a main task of achieving price stability. They do this by issuing various forms of money, fixing an array of interest rates, generating fiscal revenues, outlining the unit of account, and influencing marginal costs of production through credit regulations, among other policies (Castillo-Martinez and Reis, 2019).

The main goal of the National Bank of Romania (NBR) is to assure and maintain price stability, with monetary policy being employed under inflation targeting regime starting August 2005. In this situation, active communication of the central bank to the public at large represents an important role, and the crucial tool that the monetary authority utilizes to this end is the Inflation Report. Among other aspects, the Report highlights the National Bank of Romania's (NBR) quarterly inflation forecast over an eight-quarter period, comprising the related uncertainties and risks. Also, it provides an examination of recent and future macroeconomic context from a monetary policy decision's perspective.

Unanticipated events such as the Coronavirus pandemic or the war in Ukraine have dire consequences on the economy. Therefore, central banks should keep inflation close to the target, not only to meet its main objective, but also to ensure the optimal function of the economy. Our study aims at forecasting the Romanian inflation rate using the Autoregressive Integrated Moving Average (ARIMA). Unlike other analyses which focus on forecasting the Consumer Price Index (CPI) through the ARIMA technique, this study comprises the actual values of the monthly inflation rate from 2018 to 2022. Furthermore, there are a few studies which consider the structural break issue and solve this in their analyses. To correct the structural break issue, we construct a dummy variable and include it in our econometric model.

2. Literature review

Inflation is considered as one of the barometer tools to examine the health of an economy. A rate of inflation which is too high will diminish the level of social welfare. Conversely, a low rate of inflation shows an economy that does not function maximally, with an impact on slowing economic growth, increased poverty, and stagnant job formation. Considering these, inflation represents a macroeconomic problem that central banks should carefully consider.

The main focus of monetary policy has generally been the maintenance of a low and stable inflation rate as defined by traditionally embraced measures, such as the Consumer Price Index (CPI). The fundamental justification for this goal is the extensive consensus, supported by various economic studies, which highlight that inflation is costly insofar since it weakens the real economic activity.

Inflation forecasting represents a key aspect that the central bank of a country should consider when implementing monetary policies. Also, this plays an important role within the Eurozone, since a significant departure from the inflation target of a country might affect the stability of the Eurosystem. Therefore, given the likelihood of continuous differences in inflation levels across euro area currencies and the consequent impact on competitiveness, examining and understanding price developments in individual countries will remain of substantial importance.

There are a number of techniques available for predicting economic time-series. However, this paper focuses on the ARIMA approach. The main findings of other authors using this method are presented below.

Okafor and Shaibu (2013) deploy an univariate Autoregressive Integrated Moving-Average (ARIMA) econometric model for the Nigerian inflation rate and examine the prediction performance of the forecasted model between 1981 and 2010. Based on different selection evaluation and diagnostic criteria, the best model is chosen for the short-run prediction of Nigerian inflation rate. They discover that ARIMA (2, 2, 3) is the optimal model to use for forecasting. They confirm that the predicted inflation equation clearly shows that expected inflation represents a significant determinant of actual inflation within the estimation period.

In order to estimate a time-series model, which encompasses monthly inflation data between January 1997 and August 2013, (Baciu, 2015) applies the Box-Jenkins methodology. The ARIMA (1, 1, 2) model is selected and the inflation rate prediction for September 2013 is made. The results of the study reveal that there is a significant difference between the actual inflation rate and the forecast made on the selected model. An ARIMA (1, 1, 2) model is also used to forecast inflation in Tanzania by using the Box-Jenkins ARRIMA approach and annual inflation data from 1966 to 2017. The results demonstrate that inflation rate in Tanzania is likely to follow an upward trend in the following decade. The study encourages policymakers to apply tight fiscal and monetary policy measures to handle inflation in Tanzania (Thabani, 2019a). A similar approach is followed to forecast inflation in the Philippines. In his analysis, (Thabani, 2019b) uses annual inflation data ranging from 1960 to 2017. He finds that an ARIMA (1, 1, 3) model is stable and acceptable for forecasting inflation level in the Philippines.

In an attempt to forecast the monthly inflation rate in Pakistan, (Salam et al., 2007) deploy different ARIMA models, and the candid one is proposed. Based on different diagnostic, selection and evaluation criteria, they select the best model for the short-term prediction. Also, Jafarian-Namin et al. (2021) focus on modelling and predicting the yearly inflation level in Iran from 1960 to 2019 using ARIMA models. Different models are examined and they find that non-seasonal ARIMA (1 ,0 ,0) is the right model to predict the inflation rate for the next years.

3. Theoretical framework, methodology, and data source

3.1. Basics of ARIMA

Autoregressive Integrated Moving-Average (ARIMA) modelling represents a particular subset of univariate modelling, in which a time-series data is defined in terms of previous values of itself (the autoregressive element) and current and lagged values of an error term called "white noise" (the moving-average element).

ARIMA methodology for predicting time-series data is essentially agnostic. Contrary to other methods, it does not consider any fundamental economic model or structural associations. It is considered that past values of the series and previous error terms include information for the objectives of forecasting. The major advantage of ARIMA prediction is that it requires data on the time-series under analysis only. Firstly, this characteristic is advantageous if predicting a significant number of time-series. Secondly, this avoids an issue that appears occasionally with multivariate models, for example, it might be possible that a consistent time-series is only available for a less period than the other series, limiting the period over which the econometric model is estimated. Thirdly, in multivariate modelling, data timeliness might be a problem.

Also, this method includes several disadvantages. For example, some of the classic model identification methods are subjective and the trustworthiness of the selected model might depend on the experience and skill of the researcher (even though this criticism regularly applies to other modelling techniques as well). Another disadvantage refers to the fact that it is not embedded within any elemental theoretical model or structural relationships and therefore, the economic significance of the chosen model is not clear. Additionally, ARIMA models are basically "backward looking" and as such, they are normally poor at forecasting turning points, unless the turning point is a return to long-term equilibrium. However, ARIMA models have demonstrated to be relatively powerful, especially when producing short-run inflation predictions (Kenny et al., 1980).

As highlighted by the authors Okafor and Shaibu (2013), the ARIMA modelling approach connects two different processes into one equation. The first characteristic defines an autoregressive process, while the second characteristic represents a moving-average. ARIMA modelling promotes an association between a particular time-series data and its own lagged values.

A p^{th} – order autoregressive expresses the dependent variable as a function of its past values, as in the following equation:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

Where Y_t represents the dependent variable predicted at time t; $Y_{t-1}, Y_{t-2}, Y_{t-3}$, are the targets at time lags t-1, t-2, …, t-p; $\phi_0, \phi_1, \dots, \phi_p$ are the coefficients to be estimated; ϵ_t defines the error term at time t.

A qth – order moving-average process defines a response variable, Y_t, as a function of the lagged values of the q error terms, as in the following equation:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

Where Y_t represents the dependent variable predicted at time t; μ displays the constant mean of the process; $\theta_0, \theta_1, \dots, \theta_q$ are the coefficients to be estimated; ϵ_t highlights the error term at time t; $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$ are the errors in the past period that are included in the response variable, Y_t.

To construct an ARIMA model, we start with an econometric equation, which contains no exogenous variables ($Y_t = \beta_t + \epsilon_t$) and add to it both the autoregressive (AR) component and the moving-average (MA) component. Therefore, considering equations (1) and (2), the ARIMA (p, d, q) model takes the following form:

$$Y_t = \beta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (3)$$

Where the ϕs and θs represent the coefficients of the autoregressive and moving-average processes; d represents the order of integration of the series.

In the ARIMA modelling and forecasting technique, we consider the following steps:

1. The first step involves collecting and assessing graphically the data to be predicted. In this regard, an extensive time series data is asked for univariate time-series forecasting. It is generally recommended that at least 50 observations to be available. Using the Box-Jenkins approach can be problematic if few observations are included. Unfortunately, although a large time-series is available, it is probable that the series incorporates a structural break, which might imply analyzing a subset of the entire dataset, or otherwise using dummy variables. Furthermore, graphically exploring the data is important. Data should be examined in levels, logarithms, and differences. These steps might reveal if there is an important seasonal pattern in the series, or if any structural breaks, outliers or data errors happen. One approach to examine the characteristics of a time-series is to plot its correlogram. However, even if the correlogram offers some indication as to whether series is stationary or not, formal tests for stationarity with known statistical properties are more appropriate.
2. The second step refers to whether series is stationary or if differencing is required. Series under analysis must be stationary before identifying an appropriate ARIMA model. With this respect, Augmented Dickey-Fuller (ADF) test is performed.
3. Once the data is concluded to be stationary, we seek to discover and estimate the optimal ARIMA model. Having identified the suitable order of differencing required to make the series stationary, the next step is to find the best form of ARIMA to model the stationary data. The traditional Box-Jenkins approach consists of a repetitive process of model identification, model estimation, as well as model evaluation. The Box-Jenkins

approach represents a quasi-formal method with model identification depending on the subjective assessment of correlograms and partial correlograms plots of the data.
4. The next step involves checking if the model is a good one by deploying tests on the parameters and residuals of the model. It examines the statistical properties of the estimated ARIMA model in observing the model adequacy.
5. The last step relates to the usage of the selected model. If the model does not violate the diagnostic tests' requirements, then it can be utilized for forecasting.

3.2. Methodology

The research technique for this study is modelled according to the Box-Jenkins methodology (Box and Jenkins, 1976), which is typically applied to short-term forecasting of time-series events. The ARIMA model deployed in this study is expressed in equation 3 from above, in the defined time-series characteristics identified as (p, d, q), where p represents the order of the autoregressive (AR) component, d is the number of the differences applied on the series in order to become stationary, and q is the order of the moving-average (MA) component. This paper focuses on analyzing different ARIMA models based on monthly inflation data in Romania and use the most suitable one for prediction.

3.3. Data source

Data used in this analysis was collected from the YCharts database, encompassing a period between September 2018 and October 2022. This means that data contains 50 variables, which meets the requirement of univariate time-series modelling. EViews econometric software is used within the entire analysis.

4. Results and discussions

4.1. Visual examination of the series

The first phase in modelling time-series is to investigate the structure of the data in order to get some preliminary information regarding stationarity of the series. Before applying formal tests, the graphs of the time-series under analysis are plotted. These plots offer initial knowledge about the possible nature of the time series.

Figures show that there might be evidence of the presence of structural break in the series. Also, they highlight that building a model for the logarithmic values is likely to be more appropriate, since the modifications in the logarithmic series show a more stable variance than the modifications in the original series. The logarithm transformation helps stabilizing the variance of the series.

To check if structural breaks exist within the series, we perform the Multiple Breakpoints Tests.

Table 1. *Multiple breakpoints tests*

Break Dates	Sequential	Repartition
1	2021M10	2020M03
2	2020M02	2021M03
3	2021M03	2021M10

Source: Authors' own computations.

Table 1 shows that there are multiple structural breaks in the series. In order to keep the same sample size of 50 observations, we construct a dummy variable, which takes the value of 0 up until 2021M9 and value of 1 from 2021M10 onwards. The dummy variable is statistically significant and after including it in the regression, the null hypothesis of the CUSUMSQ test (H_0: parameters are stable) is not rejected (the blue line lies within the red lines). Therefore, the problem of structural breaks is eliminated and we can proceed with testing the stationarity of the series.

4.2. Unit root test

Series must be stationary before it can be used to identify and estimate a model. With this respect, the *Augmented Dickey-Fuller Test (ADF)* help us identifying if data is stationary or not. The test results are presented in Table 2 below.

Table 2. *Augmented Dickey-Fuller test*

Variable	Trend Specification		Remark
	Level	First Difference	
LINFL	-1.345034	-4.584306***	I(1)

Note: *, **, *** denote statistical significance at 10%, 5%, and 1% levels.
Source: Authors' own computations.

Results presented in Table 2 indicate that the ADF test statistic for the variable of interest is greater than the corresponding 95% critical values. Inflation variable became stationary after its first difference.

4.3. Model identification, estimation, and interpretation

The main objective of this paper is to assess inflation dynamics with ARIMA modelling technique. Since the logarithm of the inflation rate variable becomes stationary after taking the first order difference, the model that we are looking at is defined by ARIMA (p, 1, q). Next, we identify the best model, estimate appropriate parameters, perform diagnostic tests, and ultimately, forecast the inflation series.

The following approach is applied in estimating the univariate ARIMA model. Firstly, inflation rate series is transformed to stabilize the variable. Secondly, tentative models are identified utilizing the autocorrelation function (ACF), together with the partial autocorrelation function (PACF), and estimated through the Ordinary Least Squares (OLS) method. Thirdly, the most performant model is selected considering the smallest values for the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), the significance of the ARMA components, and the value of adjusted R^2. Fourthly, the selected model is estimated and diagnostic tests are performed. Ultimately, the estimated model is used to forecast inflation and the prediction performance is examined.

4.3.1. ARIMA model identification

We perform the series correlogram for the first difference, which consists of the ACF and PACF values. We notice the patterns in the ACF and PACF, and identify the values for the parameters p and q for the ARIMA model. The autocorrelation part from the correlogram defines the MA(q) component, while the partial autocorrelation part highlights the AR(p) component. Also, the parsimony characteristic is important, which means that including more variables in the model will raise the model fit (R^2), but at the cost of decreasing degrees of freedom. From the autocorrelation part, we select MA(1) and MA(7), while from the partial correlation part, AR(1) and AR(7) might be suitable. Therefore, there are four tentative models: ARIMA (1, 1, 1), ARIMA (7, 1, 1), ARIMA (1, 1, 7), and ARIMA (7, 1, 7).

4.3.2. ARIMA model estimation

After identifying the candidate models, it is a good practice to estimate each of them, then compare and select the most appropriate one.

Table 3. *The comparison of the candidate models*

Selection Criteria	Model			
	ARIMA (1, 1, 1)	ARIMA (1, 1, 7)	ARIMA (7, 1, 1)	ARIMA (7, 1, 7)
Akaike Information Criterion (AIC)	-1.333231	-1.370237	-1.397255	-1.242131
Schwarz Criterion (SIC)	-1.178797	-1.215803	-1.242821	-1.087697
Hannan-Quinn Criterion (HQ)	-1.274639	-1.311645	-1.338663	-1.183539
Adjusted R^2	0.183632	0.222027	0.241532	0.121301
Significance of the ARMA components	p[AR(1)]=0.8138; p[MA(1)]=0.1118	p[AR(1)]=0.0495; p[MA(7)]=0.0624	p[AR(7)]=0.0512**; p[MA(1)]=0.0042*	p[AR(7)]=0.9971; p[MA(7)]=0.5822

Note: *, **, ** denote statistical significance at 10%, 5%, 1% levels.
Source: authors' own computations.

The results in Table 3 indicate that ARIMA (7, 1, 1) is the appropriate one, since it has the lowest AIC, SIC, HQ, the ARMA components are significant, and adjusted R^2 is the highest. Therefore, it is reasonable to proceed with this model in our analysis.

4.3.3. Diagnostic tests

It is necessary to analyse the statistical properties of the estimated ARIMA model in observing the model adequacy. The model is tested for specification error, normality of the residuals, serial correlation, and heteroskedasticity. Additional checks will be made in order to see if the estimated AR(I)MA process is (covariance) stationary (if AR roots lie inside the unit circle) and if the estimated AR(I)MA process is invertible (if all MA roots lie inside the unit circle).

Table 4. *ARIMA (7, 1, 1) Diagnostic tests*

Test	F-statistic	P-value
Jarque-Berra test	0.874403	0.645841
Ramsey RESET test	0.104598	0.7479
Breusch-Godfrey LM test	0.097736	0.9071
Autoregressive Conditional Heteroskedasticity (ARCH) test	1.035766	0.3141

Source: Authors' own computations.

Table 5. *Stationarity of the AR(I)MA process*

AR Root(s)	Modulus
-0.183580 $\pm$ 0.804316i	0.825000
-0.743299 $\pm$ 0.357954i	0.825000
0.514379 $\pm$ 0.645011i	0.825000
0.825000	0.825000
Conclusion: No root lie outside the unit circle	

Source: Authors' own computations.

Table 6. ***Invertibility of the AR(I)MA process***

MA Root(s)	Modulus
-0.440065	0.440065
Conclusion: No root lie outside the unit circle	

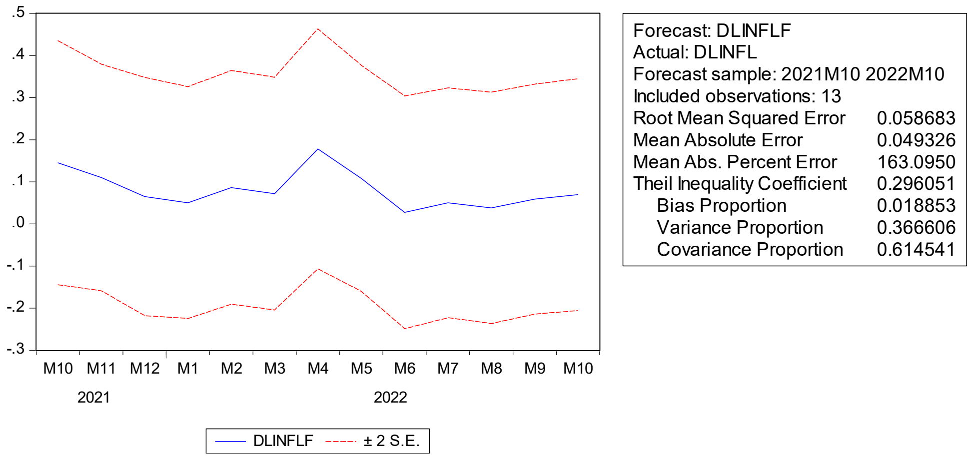
Source: Authors' own computations.

Results in Table 4 indicate that the model is well specified based on the Ramsey RESET test and serially uncorrelated on the basis of the Breusch-Godfrey serial correlation LM test. The Autoregressive Conditional Heteroskedasticity (ARCH) test shows that there is no fluctuation clustering in Romania's inflation monthly data. Also, the Jarque-Berra (JB) test that the residuals are normally distributed.

4.3.4. Forecast evaluation for ARIMA (7, 1, 1) model

Next, we forecast the Romanian inflation rate using the ARIMA (7, 1, 1) model. The period of forecasts is from October 2021 to October 2022.

Figure 1. *The forecast of the inflation rate for the period October 2021 – October 2022*



In Figure 1, the blue line is the forecast value of inflation. The red lines which are above and below the predicted monthly inflation rate display the forecast with ± 2 of standard errors. Some prediction measurements such as the mean absolute error (MAE), root mean squared error (RMSE), and Theil inequality coefficient are presented. The estimated result for ARIMA in Figure 1 shows that the model is a good one.

5. Conclusion

In this paper, we present the ARIMA model and the steps required in order to make forecasts for a particular series. We discover that ARIMA (7, 1, 1) is the best one and hence, we use this model to forecast the monthly inflation rate in Romania. However, for a higher increase in the accuracy of the forecasts, we suggest two aspects to be considered for future research. The first one refers to using models from machine learning/deep learning area. Even if some of them are very complex and difficult to interpret, they ensure high performance of the models. There will be a trade-off between model efficiency and interpretability. The second suggestion relates to an expansion of the sample size. This is also a pre-requisite for using complex algorithms, since these require hundreds even thousands of observations, depending on the algorithm type. We envisage that machine learning models will replace many traditional algorithms, which are currently used in the central bank area. However, we do not exclude the possibility of a mix use of traditional and machine learning models.

References

Baciu, I.C., 2015. Stochastic models for forecasting inflation rate. Empirical evidence. *Procedia Economics and Finance,* Vol. 20, pp. 44-52. doi: 10.1016/S2212-5671(15)00045-3

Box, G. and Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control, s.l.: San Francisco: Holden-Day Inc.

Castillo-Martinez, L. and Reis, R., 2019. How do central banks control inflation? A guide for the perplexed, s.l.: LSE Manuscript.

Csorba, L.M. and Juravle, D.J., 2022. Study Regarding the Impact of the COVID-19 Pandemic on Inflation in Romania in the 2019-2020 period. *Revista Economica,* 74(1), pp. 51-64.

Jafarian-Namin, S., Fatemi-Ghomi, S., Shojaie, M. and Shavvalpour, S., 2021. Annual forecasting of inflation rate in Iran: Autoregressive integrated moving average modeling approach. *Engineering Reports,* 3(4), pp. 1-11. doi: 10.1002/eng2.12344

Kenny, G., Meyler, A. and Quinn, T., 1998. Forecasting Irish inflation using ARIMA models, s.l.: Central Bank of Ireland (No. 3/RT/98).

Majumder, S.C., 2016. Inflation and Its Impacts on Economic Growth of Bangladesh. *American Journal of Marketing Research,* 2(1), pp. 17-26.

Okafor, C. and Shaibu, I., 2013. Application of ARIMA Models to Nigerian Inflation Dynamics. *Research Journal of Finance and Accounting,* 4(3), pp. 138-150.

Salam, M.A., Salam, S. and Feridun, M., 2007. Modeling and forecasting Pakistan's inflation by using time series ARIMA models, s.l.: Economic Analysis Working Papers (No. 2007, 1).

Thabani, N., 2019a. Modeling and forecasting inflation in Tanzania using ARIMA models, s.l.: Munich Personal RePEc Archive (MPRA No. 92458).

Thabani, N., 2019b. Modeling and forecasting inflation in Philippines using ARIMA models, s.l.: Munich Personal RePEc Archive (MPRA No. 92429).

Yolanda, Y., 2017. Analysis of Factors Affecting Inflation and its Impact on Human Development Index and Poverty in Indonesia. *European Research Studies Journal,* XX(4B), pp. 38-56. doi: 10.35808/ERSJ/873