ARIMA model forecast based on EViews software

Lihua Ma*, Chao Hu, Rongchao Lin and Yanben Han
National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China.
Corresponding author e-mail: mlh@nao.cas.cn

Abstract. Time series is a series of data obtained in chronological order. Future values of most time series can be forecasted according to current values and past values. The EViews software is a software package specifically designed to process time series data. Autoregressive Integrated Moving Average (ARIMA) model, a time series forecast method, can be achieved with the EViews software. Based on the EViews software, the forecast procedure with ARIMA model is illustrated in this work. As an example, the Gross Domestic Product (GDP) of China is forecasted from 2016 to 2018.

1. Introduction
EViews is an acronym for Econometrics Views. Literally, it is translated as econometric observations, often referred to as econometric software packages. Originally, it can be used to “observe” quantitative rules between socio-economic relations and economic activities, with econometrics method and technique. The software is a tool package developed by Quantitative Micro Software (QMS) of the United States for data analysis, regression analysis and prediction under the Windows operating system. It can be used to quickly find statistical relationships from data and predict future values. EViews combines spreadsheets, database technologies and traditional statistical software analysis capabilities and provides advantage of the visualization operations for modern Windows software. It uses the mouse to operate standard Windows menus and dialogs, and the related results appear in the window and can be processed using standard Windows technology. Additionally, EViews has powerful command functions and batch language features. Some commands including edit and execute can be inputted on the command line and stored in program files which can easily be called in subsequent projects. It supports some file format such as Excel, SPSS, SAS, Stata, RATS, TSP. It can connected to the ODBC database. EViews has been widely used in the fields of applied in econometrics, macroeconomic forecasting, sales forecasting, financial analysis, cost analysis, Monte Carlo simulation, data analysis and evaluation. It has become one of the most widely used econometric statistical software in the world.

With the development of society, the deepening of theoretical research and the improvement of observation methods, people have acquired more and more data series. Time series analysis plays an increasingly important role in statistics and forecasting techniques. As for stationary time series, someone can use the Autoregressive (AR) model, the Moving Average (MA) model, or the Autoregressive and Moving Average (ARMA) model to fit series and forecast the future value of the time series. Actually, some time series are often non-stationary. Someone has to make a differential operation. Then ARMA model prediction can be applied. The process is so-called the Autoregressive Integrated Moving Average Model (ARIMA) [1, 2]. Based on the EViews software, the modeling and forecast procedure with ARIMA model is illustrated in this work.
2. ARIMA Model

The ARIMA model, a time series prediction method, was proposed by Box and Jenkins in the 1970s. The model consists of AR, I, and MA. Here AR represents the Autoregressive model, I represents the Integration indicating the order of single integer, and MA represents the Moving Average model. In general, a stationary sequence can establish a metrology model. The unit root test is used to judge the stationarity of the sequence. As for a non-stationary sequence, it should be converted to a stationary sequence with difference operation. The number of corresponding difference is called as the order of single integer. The ARIMA (p, D, q) model is essentially a combination of differential operation and ARMA (p, q) model [3, 4]. A non-stationary I (D) process is one that can be made stationary by taking D differences. The process is often called difference-stationary or unit root processes.

A series that can be modeled as a stationary ARMA (p,q) process after being differenced D times is denoted by ARIMA (p,D,q) [5]. The form of the ARIMA (p,D,q) model is

\[
\Delta^D y_t = c + \phi_1 \Delta^D y_{t-1} + \ldots + \phi_p \Delta^D y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}. \tag{1}
\]

Where \(\Delta^D y_t\) denotes a D-th differenced series, and \(\epsilon_t\) is an uncorrelated process with mean zero. In lag operator notation, \(L^i y_t = y_{t-i}\). The ARIMA (p,D,q) model can be written as

\[
\phi(L) y_t = \theta(L) \epsilon_t. \tag{2}
\]

Here, \(\phi(L)\) is an unstable AR operator polynomial with exactly D unit roots. Someone can factor this polynomial as \(\phi(L)(1-L)^D\), where \(\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p\) is a stable degree p AR lag operator polynomial. Similarly, \(\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q\) is an invertible degree q MA lag operator polynomial.

When two out of the three terms in ARIMA(p,D,q) are zeros, the model may be referred to, based on the non-zero parameter, dropping “AR”, “I” or “MA” from the acronym describing the model. For example, ARIMA (1,0,0) is AR (1), ARIMA (0,1,0) is I (1), and ARIMA (0,0,1) is MA (1).

The ARIMA model is a commonly used time series model and a short-term prediction model with high precision. The basic idea of the model is that some time series are a set of random variables that depend on time, but the changes of the entire time series have certain rules, which can be approximated by the corresponding mathematical model. Through the analysis of the mathematical model, it can understand the structure and characteristics of time series more fundamentally and achieve the optimal prediction in the sense of minimum variance.

3. ARIMA modeling and forecasting

3.1. Procedure of ARIMA modeling

The procedure flow chart of ARIMA modeling and forecasting is given in Figure 1.

![Figure 1. The procedure flow chart of ARIMA modeling and forecasting.](image-url)
(1) Identifying the stationarity of the time series. The stationarity of the sequence is judged based on line graph, scatter plot, autocorrelation function and partial autocorrelation function graphs of the time series. The unit root of Augmented Dickey-Fuller (ADF) is usually used to test the variance, trend and seasonal variation and identified the stationarity.

(2) Determining the order of single integer D. If the time series is a stationary sequence, going directly into Step (3). If the time series is a non-stationary sequence, appropriate transformation (including difference, variance stationarity, logarithm, square root) should be used to be converted to a stationary sequence. The number of differences is the order of single integer.

(3) ARMA modeling. As for the result sequence of Step (2), autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) of the sequence are calculated. And the values of the autocorrelation order p and the moving average order q of the ARMA model can be estimated. The basic principle for determining the order p and q is given in Table 1.

| Autocorrelation coefficient (ACF) | Partial autocorrelation coefficient (PACF) | Model order |
|----------------------------------|------------------------------------------|-------------|
| / p-order truncation             | / q-order truncation                      | AR(p)       |
| trailing                         | trailing                                  | MA(p,q)     |
|                                  |                                          | ARMA(p,q)   |

(4) Performing parameter estimation. The autocorrelation and partial autocorrelation graphs are used to judge the number of autocorrelation coefficients and partial autocorrelation coefficients with remarkable significant level. In the step the roughly model of the sequence can be selected.

(5) Diagnostic test and optimization. The model is diagnosed and optimized by performing a white noise test on the residual. If the residual is not a white noise, return to Step (4) and re-select the model. If the residual is a white noise, return to Step (4) and create multiple models, and choose the optimal model from all the fitted model of the test.

### 3.2. Procedure of ARIMA forecasting

The future value of a time series can be forecasted with the ARIMA model. An important application of EViews software is modeling and prediction based on ARIMA model. The EViews software gives two prediction methods, Static and Dynamic. Static is a one-step advance prediction, and Dynamic is a short-term dynamic prediction. The procedure is as follows:

1. If the time series is a non-stationary sequence, it should be firstly converted to a stationary sequence. The best model parameters are selected and the ARIMA (p, D, q) model is established.

2. In the Equation window of the EViews software, select the Forecast menu. In the dialog box, Static or Dynamic can be selected as needed. Someone can modify the name of forecasting sequence or use the default value, and click OK.

### 4. Modeling and forecasting example based on ARIMA model

#### 4.1. Data description

The Gross Domestic Product (GDP) is the core indicator of national economic accounting. It is an important index to measure the overall economic situation of some country. It reflects the country’s economic strength, structural layout and market scale. In 2016, the National Bureau of Statistics of China revised research and development expenditure accounting method according to the international standard of national economic accounting jointly promulgated by the five major international organizations, the National Account System 2008 was revised. The revised China GDP data from 1952 to 2015 is listed in Table 2. We make modeling and prediction of the GDP data in the following sections.
Table 2. China GDP data from 1952 to 2015.

| Year | GDP (hundred million RMB) | Year | GDP (hundred million RMB) | Year | GDP (hundred million RMB) |
|------|---------------------------|------|---------------------------|------|---------------------------|
| 1952 | 679                        | 1974 | 2828                      | 1996 | 71814                     |
| 1953 | 824                        | 1975 | 3039                      | 1997 | 79715                     |
| 1954 | 860                        | 1976 | 2989                      | 1998 | 85196                     |
| 1955 | 912                        | 1977 | 3250                      | 1999 | 90564                     |
| 1956 | 1031                       | 1978 | 3679                      | 2000 | 100280                    |
| 1957 | 1071                       | 1979 | 4100                      | 2001 | 110863                    |
| 1958 | 1312                       | 1980 | 4588                      | 2002 | 121717                    |
| 1959 | 1448                       | 1981 | 4936                      | 2003 | 137422                    |
| 1960 | 1470                       | 1982 | 5373                      | 2004 | 161840                    |
| 1961 | 1232                       | 1983 | 6021                      | 2005 | 187319                    |
| 1962 | 1162                       | 1984 | 7279                      | 2006 | 219438                    |
| 1963 | 1248                       | 1985 | 9099                      | 2007 | 270232                    |
| 1964 | 1470                       | 1986 | 10376                     | 2008 | 319516                    |
| 1965 | 1734                       | 1987 | 12175                     | 2009 | 349081                    |
| 1966 | 1889                       | 1988 | 15180                     | 2010 | 413030                    |
| 1967 | 1794                       | 1989 | 17180                     | 2011 | 489301                    |
| 1968 | 1744                       | 1990 | 18873                     | 2012 | 540367                    |
| 1969 | 1962                       | 1991 | 22006                     | 2013 | 595244                    |
| 1970 | 2280                       | 1992 | 27195                     | 2014 | 643974                    |
| 1971 | 2457                       | 1993 | 35673                     | 2015 | 685506                    |
| 1972 | 2552                       | 1994 | 48637                     |      |                           |
| 1973 | 2756                       | 1995 | 61340                     |      |                           |

4.2. Stationarity test

The GDP data series during 1952-2015 is plotted in Figure 2. The result of the stationarity test (ADF test) on the data is given in Table 3.

![GDP graph](image)

Figure 2. The GDP data during 1952 to 2015.

Table 3. Augmented Dickey-Fuller unit root test on GDP.

| Augmented Dickey-Fuller test statistic | 3.651882 | 1.0000 |
| Test critical values:                  |          |        |
| 1% level                               | -3.552666|        |
| 5% level                               | -2.914517|        |
| 10% level                              | -2.595033|        |
It can be seen that ADF=3.651882 is greater than the critical value of the significance level of 0.01, 0.05 and 0.1, that is to say, the original GDP sequence is non-stationary.

In Figure 2, the original sequence is exponential. Taking the natural logarithm of the GDP data to eliminate its non-stationary and obtaining the LGDP sequence. And taking LGDP for ADF test, ADF=1.397117 is still greater than the critical value of the significance level of 0.01, 0.05 and 0.1. The LGDP sequence still accepts the null hypothesis with a large P value. The LGDP sequence is still non-stationary. Further, the first-order difference is performed and a DLGDP sequence is obtained. The results of the ADF test for the DLGDP sequence is given in Table 4.

**Table 4. Augmented Dickey-Fuller unit root test on DLGDP.**

| Augmented Dickey-Fuller test statistic | t-Statistic | Prob.* |
|----------------------------------------|-------------|--------|
| Augmented Dickey-Fuller test statistic | -4.340237   | 0.0009 |
| Test critical values:                  |             |        |
| 1% level                               | -3.542097   |        |
| 5% level                               | -2.910019   |        |
| 10% level                              | -2.592645   |        |

It can be seen that ADF=-4.340237 is less than the three critical values of the test level. That is to say, the DLGDP sequence after the logarithmic change and the first-order difference is a stationary series, and the significance test of the stationarity is passed. It can be seen that the original GDP sequence is a first-order single-order sequence, that is, LGDP ~ I (1).

### 4.3. Model Identification

With the EViews software, the autocorrelation and partial autocorrelation function graphs of the DLGDP series are plotted in Figure 3.

**Figure 3.** Autocorrelation and partial autocorrelation function graphs of the DLGDP series.

It can be seen from Figure 3 that the autocorrelation coefficient of the DLGDP sequence is significantly non-zero when the lag order is 1. And it is basically in the confidence band when the lag order is greater than 1, so q can be taken 1. The partial autocorrelation coefficient is significantly non-zero when the lag order is equal to 1, and it is also significantly different from 0 when the lag order is 2, so p=1 or p=2 can be considered. Considering that the judgment is very subjective, to establish a more accurate model, the range of values of p and q is appropriately relaxed, and multiple ARMA (p, q)
models are established. The order with 0, 1, 2 in autoregressive moving average pre-estimation is performed on the processed sample data. Table 5 lists the test results of ARMA (p, q) for different parameters. Adjusted R-squared, AIC value, SC value and S.E. of regression are all important criteria for selecting models. The AIC criterion and the SC criterion are mainly used for ranking, and select the optimal model. Generally, the larger the coefficient of determination, the smaller the AIC value and the SC value, and the residual variance. The corresponding ARMA (p,q) model is superior.

| (p,q)   | Adjusted R-squared | AIC       | SC        | S.E. of regression |
|--------|--------------------|-----------|-----------|-------------------|
| (0,1)  | 0.312800           | -2.520025 | -2.417971 | 0.066851          |
| (0,2)* | 0.016962           | -2.166110 | -2.064056 | 0.079956          |
| (1,0)  | 0.330593           | -2.545679 | -2.443625 | 0.065980          |
| (1,1)* | 0.353535           | -2.564026 | -2.427954 | 0.064839          |
| (1,2)  | 0.338601           | -2.542001 | -2.405929 | 0.065584          |
| (2,0)* | 0.010211           | -2.160068 | -2.058014 | 0.080230          |
| (2,1)  | 0.336333           | -2.538841 | -2.402769 | 0.065696          |
| (2,2)  | 0.153379           | -2.236239 | -2.100167 | 0.074201          |

It should be emphasized that although the appropriate ARMA model is usually selected using the AIC value and the SC value. However the minimum AIC value and the SC value are not sufficient conditions for the optimal ARMA model. The method used in this work is to first establish a model with the minimum AIC value and SC value, and perform a parameter significance test and a residual randomness test on the estimation result. If it passes the test, the model can be regarded as the optimal model; if it cannot pass the test, the second smallest AIC value and SC value are selected and the relevant statistical test is performed. And so on, until the appropriate model is selected. In Table 5, the model that did not pass the parameter significance test and the residual randomness test was identified by “*”. Finally, it is preferable to prefer the ARMA (1, 0) model.

4.4. Model establishment and inspection

The estimated results with the ARIMA model are as follows:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|-------|
| C        | 0.110597    | 0.020849   | 5.304763    | 0.0000|
| AR(1)    | 0.590915    | 0.113374   | 5.212098    | 0.0000|
| SIGMASQ  | 0.004146    | 0.000679   | 6.102246    | 0.0000|

R-squared 0.352187
Adjusted R-squared 0.330593
S.E. of regression 0.065980
Sum squared resid 0.261198
Log likelihood 83.18890
F-statistic 16.30964
Prob(F-statistic) 0.00002

Inverted AR Roots .59

The final model of the LGDP sequence is ARIMA (1, 1, 0), and Equation (3) displays the specific form of the model. Here the data in parentheses below the equation is the t-test statistic of the corresponding estimate value.
The estimated value of the variance of the corresponding error term is

$\sigma_a^2 = 0.065980$ (4)

It can be seen from the t statistic of the model coefficients and its P value that the parameter estimates of all explanatory variables of the model are significant at the significance level of 0.01.

The model is used to fit the DLGDP data, and the result is shown in Figure 4. In the figure, the actual data is given by the solid line, and the upper and lower dotted lines correspond to the fitted values and residual of the model.

![Figure 4](image_url)

**Figure 4.** Actual series, fitted series and residual series of the DLGDP sequence.

A white noise test is performed on the residual after fitting the ARIMA (1, 1, 0) model. The autocorrelation and partial autocorrelation function graphs of the residual series are shown in Figure 5. It can be seen that the residual is a white noise, indicating that the model is valid.

![Figure 5](image_url)

**Figure 5.** Autocorrelation and partial autocorrelation function graphs of the residual series.
4.5. Data forecasting
Firstly, the model is used to analyse the fitting effect with the GDP value in 2015. The forecast value in 2015 is 705847.6 hundred million RMB. The actual value is 685506 hundred million RMB and the relative error is 2.97%. It can be seen that the forecast value is close to the actual result, indicating that the model has a good fitting effect.

Under the graphical interface in EViews software, Dynamic forecast mode is used to predict the GDP values from 2016 to 2018. The results are listed in Table 7.

The National Bureau of Statistics of China has not officially released revised GDP data in 2016 and 2017. According to the official website of the National Bureau of Statistics, the verification value of GDP in 2016 is 743585 hundred million RMB, and the preliminary value of GDP in 2017 is 827122 hundred million RMB. Here, the relative errors of GDP forecasting in 2016 and 2017 are 0.09% and 1.17%, respectively.

Table 7. China GDP forecast from 2016 to 2018.

| Year  | 2016    | 2017    | 2018    |
|-------|---------|---------|---------|
|       | Forecast GDP (hundred million RMB) | 744216.9 | 817405.2 | 903980.5 |
|       | Release GDP (hundred million RMB)  | 743855   | 827122   | /        |
|       | Relative error (%)                 | 0.09     | 1.17     | /        |

5. Summary
ARIMA model forecast is a relatively advanced time series prediction method. It can realistically describe the dynamic change rules. It can be used to perform statistical analysis and forecast for time series under certain conditions. Specially, the model is suitable for short-term predictions. Large deviations occur when the forecasting time scale is long. Based on EViews software, this work gives time series modeling and forecasting with the ARIMA model. It should be noted that as for a specific time series that is subject to many factors, model predictions that rely solely on current values and historical data sometimes have a certain degree of deviation from the real situation.

Acknowledgments
This work was financially supported by National Natural Science Foundation of China (11573041, 11473045) fund.

References
[1] X.T. Zhang. A Guide to Using EViews. China Machine Press, Beijing, 2007.
[2] D.W. Zhang, B. Liu, Q. Liu. Eviews Data Statistics and Analysis Tutorial. Tsinghua University Press, Beijing, 2010.
[3] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung. Time Series Analysis: Forecasting and Control. 5th ed. Wiley, New York, 2015.
[4] H.H. Fan, L.Y. Zhang. EViews Statistical Analysis and Application. China Machine Press, Beijing, 2009.
[5] Information on https://ww2.mathworks.cn/help/econ/arima-model.html
[6] L. Li. Application research of EViews software in ARIMA model. Journal of Anhui Vocational College of Electronics & Information Technology, 53 (2011) 31-32, 51.
[7] D.M. Xue. Application of the ARIMA model in time series analysis. Journal of Jilin Institute of Chemical Technology. 27 (2010) 80-83.
[8] C.C. Zhao, Z.Y. Shang. Application of ARMA Model on prediction of Per Capita GDP in Chengdu City. Ludong University Journal (Natural Science Edition). 28 (2012) 223-226.
[9] L. Zhang. Time series model and forecast of GDP per capita in Tianjin. Northern Economy. 3 (2007) 44-46.