ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series?

Ďurka Peter[1], Pastoreková Silvia[2]

Abstract. Nowadays, there are a lot of methods and techniques to analyze and forecast time series. One of the most used is methodology based on autoregressive integrated moving average (ARIMA) model by Box and Jenkins [1]. This method uses historical data of univariate time series to analyze its own trend and forecast future cycle.

Time series are often affected by special events such as legislative activities, policy changes, environmental regulations, and similar events, which we shall refer to as intervention events. You can incorporate one or more time series in a model to predict the value of another series, by using a transfer function. Transfer functions can be used both to model and forecast the response series and to analyze the impact of the intervention.

The general transfer function model employed by the ARIMA procedure was discussed by Box and Tiao [2]. When an ARIMA model includes other time series as input variables, the model is sometimes referred to as an ARIMAX model. Pankratz [4] refers to the ARIMAX model as dynamic regression.

In this article, we use both ARIMA and ARIMAX approaches to analyze and forecast macroeconomic time series and decide whether more complex ARIMAX model brings so much better results than simple ARIMA model.

Keywords: ARIMA, transfer function model, TFM, ARIMAX, gross domestic product per capita, forecast.

JEL Classification: C22, C53
AMS Classification: 91B84

1 ARIMA model

An „AutoRegressive Integrated Moving-Average" (ARIMA) model belongs to the one of the most used methodology approaches for analyzing time series. This is mostly because of it offers great flexibility in analyzing various time series and because of achieving accurate forecasts, too. Its other advantage is that for analyzing single time series it uses its own historical data.

The ARIMA model methodology was first introduced by Box and Jenkins in 1976 [1], and ARIMA models are often referred to as Box-Jenkins models. This approach analyzes univariate stochastic time series, i. e. error term of this time series. For this to be possible, the analyzed time series must be stationary. This means that the mean, variance and covariance of the series are all constant over time. However, most economic and financial time series show trends over time. Stationarity is important because, if the series is non-stationary, all the typical results of the classical regression analysis are not valid. Regressions with non-stationary series may have no meaning and are therefore called „spurious". Long-term forecasts of a stationary series will converge to the unconditional mean of the series.

ARIMA model (with seasonal terms) can be written as follows:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + ... + \Phi_P y_{t-Ps} + \\ + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q} - \Theta_1 a_{t-s} - \Theta_2 a_{t-2s} - ... - \Theta_Q a_{t-Qs} \quad (1)$$

Using backshift (lag) operator we can rewrite (1):

$$\phi_p(B)\Phi_P(B^s)z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (2)$$

where:

$z_t = (1-B)^d(1-B^s)^D ln(y_t)$

$\phi_p(B)$ – nonseasonal operator of autoregressive process $AR(p)$

1 University of Economics, Department of statistics, Dolnozemská cesta 1, 852 35 Bratislava 5, Slovakia e-mail: peter.durka@euba.sk

2 University of Economics, Department of statistics, Dolnozemská cesta 1, 852 35 Bratislava 5, Slovakia e-mail: pastorekova.silvia@gmail.com

$\theta_q(B)$ – nonseasonal operator of moving average $MA(q)$
$\Phi_P(B^s)$ – seasonal operator of autoregressive process $AR(P)$
$\Theta_Q(B^s)$ – seasonal operator of moving average $MA(Q)$
a_t – error term (white noise)
s – orders of season ($B^s y_t = y_{t-s}$)
d, D – nonseasonal and seasonal orders of differencing (integration)

Then, using more parsimonious notation, we can rewrite (2) as follows:

$$\text{ARIMA}\ (p, d, q)(P, D, Q)_s, \tag{3}$$

where: p, P – number of autoregressive parameters
q, Q – number of moving average parameters

The Box-Jenkins approach is iterative three-stage modeling approach – identification, estimation and diagnostic checking, and finally forecasting.

In the identification stage, the researcher visually examines the time plot of the series autocorrelation function (ACF) and partial autocorrelation function (PACF). Plotting each observation of the series against time t provides useful information concerning outliers, missing values and structural breaks in the data. The analyzed time series must be stationary. Once stationarity has been achieved (logarithm and/or differences), the next step is to identify the parameters of the model, i. e. AR and MA orders examining ACF and PACF.

In the estimation stage, each of the tentative models is estimated and the various coefficients are examined. The estimated models are compared using the Akaike information criterion and the Schwarz Bayesian criterion and model with the smallest criterion is chosen to get the parsimonious model. The main approaches to fitting Box–Jenkins models are non-linear least squares and maximum likelihood estimation.

In the diagnostic checking stage, the goodness of fit of the model is examined. Residuals should meet white noise assumptions, i. e. autocorrelation, homoskedasticity and normality is tested. If these assumptions are not satisfied, one needs to fit a more appropriate model. Care must be taken here to avoid overfitting.

The main function of ARIMA models is forecasting. Their forecasting ability can be considered when compared to actual time series.

2 ARIMAX – transfer function model

Assume two time series denoted Y_t and X_t, which are both stationary. Then, the *transfer function model* (TFM) can be written as follows:

$$Y_t = C + \nu(B)X_t + N_t \tag{4}$$

where:
Y_t is the output series (dependent variable),
X_t is the input series (independent variable),
C is constant term,
N_t is the stochastic disturbance, i.e. the noise series of the system that is independent of the input series.
$\nu(B)X_t$ is the transfer function (or impulse response function), which allows X to influence Y via a distributed lag.
B is backshift operator, thus we can write

$$\nu(B)X_t = (\nu_0 + \nu_1 B + \nu_2 B^2 + \ldots)X_t, \tag{5}$$

When X_t and N_t are assumed to follow ARMA model, equation (4) is known as the ARMAX model. This ARMAX model is quite different from ARMA model, because we work with two different series X_t and Y_t - output series Y_t is related to input series X_t.

Coefficients ν_j are called impulse response weights, which could be positive or negative. The larger the absolute value of any weight ν_j is, the larger is the response of Y_t to a change in X_{t-j}. Output series might not react immediately to a change in input series, thus some initial ν weights may be equal to zero. The number of ν weights equal to zero is called dead time and is denoted as b (Rublikova, Marek [5]).

Theoretically, the transfer function $\nu(B)X_t$ has an infinite number of coefficients. Then, we can write transfer function as the rational polynomial distributed lag model of finite order as the ratio of a low order polynomials in B:

$$\nu(B)X_t = \frac{\omega_h(B)B^b}{\delta_r(B)} X_t \tag{6}$$

where $\omega_h(B) = \omega_0 + \omega_1 B + \ldots + \omega_h B^h$; $\delta_r(B) = 1 - \delta_1 B - \ldots - \delta_r B^r$; h is the number of terms plus one of the independent variable included; r is the number of terms of the dependent variable included and b is dead time mentioned above already.

Disturbance series N_t can be written in the form of an autoregressive integrated moving average model as follows:

$$N_t = \frac{\theta(B)\Theta(B^S)}{\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D} a_t \qquad (7)$$

where a_t is zero mean and normally distributed white noise.

Then, substitute (5) with maximum lag denoted by K (free-form distributed lag model) and (7) into (4), we have transfer function model in its full formula:

$$Y_t = C + \nu_0 X_t + \nu_1 X_{t-1} + \nu_2 X_{t-2} + \cdots + \nu_K X_{t-K} + \frac{\theta(B)\Theta(B^S)}{\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D} a_t \qquad (8)$$

Construction of TFM is similar iterative process as construction of univariate Box-Jenkins ARIMA model, i.e. identification, estimation and diagnostic chcecking. After checking there is no feedback from earlier values of the output to current values of the input, we can start with the linear transfer identification method (LTF) to find out the orders (*b, r, h*) of a rational form transfer function (Pankratz [4]). First, we specify free-form distributed lag model in which K is chosen according to the analyst judgment and then we specify low order for disturbance series N_t. Nonlinear least square method can be used to estimate parameters. After estimation of the model, we have to check estimated disturbance series for stationarity by means of sample autocorrelation function and sample partial autocorrelation function. If the disturbance series is not stationary, then it is necessary to difference input and output accordingly. If the disturbance is stationary, then we are going to the stage 2 where we may use preliminary estimated impulse response weights to choose the orders (*b, r, h*) of one/few tentative rational form transfer function(s) to represent $\nu(B)$. We can identify the orders (*b, r, h*) by visually comparing the estimated impulse response function with some common theoretical functions. If the linear transfer function model is adequate then we can compute forecasts. There are several diagnostic checks to decide whether the model is adequate based on the residuals which should be independent as well as input series, e.g. cross-correlation check and/or autocorrelation check.

It is good practice to build an ARIMA model for both the output and the input series before attempting to build a transfer function model (Rublikova, Marek [5])

3 Transfer function model for gross domestic product per capita

In this applied part, we are going to build a transfer function model (TFM) for gross domestic product per capita and unemployment rate. We assume that a change in unemployment rate will affect trend in gross domestic product per capita which will lead to a significant change.

3.1 ARIMA model

First, we are going to find best fitted ARIMA model for output and input series. Output series is gross domestic product per capita (GDPpc). Analyzed quarterly data cover the period from 2000 up to 2011 in thousand EUR constant prices. The source of data is Slovstat database. The plotted series is shown at Fig. 1. Because the series is nonstationary, we use difference and seasonally difference to reach the stationarity. After study of ACF and PACF of stationary series, we identify the right model as ARIMA (0,1,0)(1,1,0) written as (standard error is in parentheses):

$$(1-B)(1-B^4)(1+\underset{(0.1441)}{0.3057}B^4)HDPpc_t = a_t \qquad (9)$$

Unemployment rate (UR) is used as input series. It covers quarterly data of period from 2000 up to 2011 in percents. The source of data is Slovstat database, too. The plotted series UR is shown at Fig. 2. The series is nonstationary and just difference is appropriate to transform to its stationarity. Then, according to its ACF and PACF we identify the fitted model as ARIMA (0,1,1)(2,0,0):

$$(1-B)(1-\underset{(0.1397)}{0.5894}B^4-\underset{(0.1381)}{0.3545}B^8)UR_t = (1+\underset{(0.1116)}{0.7235}B)a_t \qquad (10)$$

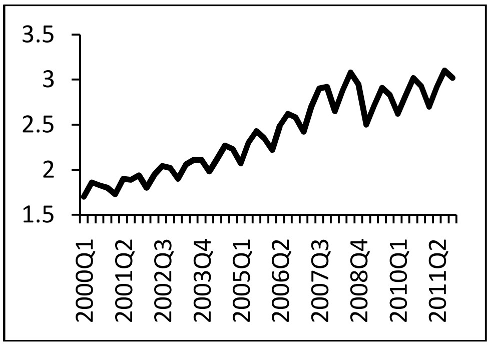

Fig. 1 Gross domestic product per capita in 1000 EUR constant prices, 2000Q1 – 2011Q4
(Source: Slovstat database)

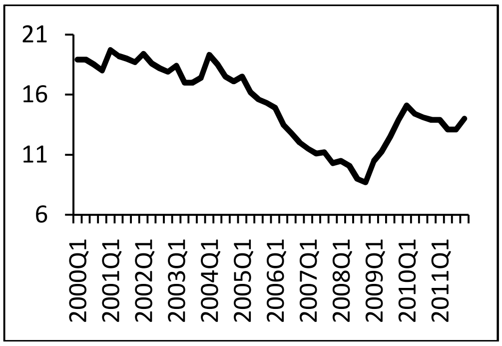

Fig. 2 Unemployment rate in %, 2000Q1 – 2011Q4
(Source: Slovstat database)

3.2 Transfer function model (ARIMAX model)

Now, we apply earlier described LTF method to build transfer function model describing relationship between GDP per capita and unemployment rate. First step is to estimate free-form distributed lag model with input and output series not differenced. We assume maximum lag $K=8$ and noise series N_t is approximated by AR(1). Estimated noise model $\hat{n}_t = y_t - \hat{y}_t$ is stationary. Only statistically significant weight was v_0. Therefore, the transfer function model has the simplest form with parameters (0, 0, 0) and transfer function is $v(B) = \omega_0$.

Next step is to build ARIMA model for noise series N_t. After long identification and estimation, we choose the model in the form ARIMA (0,0,1)(0,0,1):

$$N_t = (1-\theta_1 B)(1-\Theta_1 B^4)a_t \tag{11}$$

Now, we can estimate transfer function model in the form

$$Y_t = C + v_0 X_t + (1-\theta_1 B)(1-\Theta_1 B^4)a_t \tag{12}$$

$$GDPpc_t = \underset{(0.2144)}{3.8983} - \underset{(0.0139)}{0.0975}UR_t + (1-\underset{(0.1000)}{0.7756}B)(1-\underset{(0.1361)}{0.6049}B^4)a_t \tag{13}$$

All parameters are statistically significant. Residuals meet white noise assumptions. To check the adequacy of the fitted transfer function model, we calculate the cross-correlation function between TFM residuals and ARIMA model for unemployment rate residuals (estimated above). No values of the cross-correlation function are statistically significant, therefore residuals are not autocorrelated and fitted transfer function model is adequate. The R-Squared statistic indicates that the model as fitted explains 92.7% of the variability in gross domestic product per capita.

3.3 Forecasts of gross domestic product per capita by ARIMA and ARIMAX model

Now, we will compute not only forecasts by fitted transfer function model (ARIMAX) above but also the forecasts for individual series of GDP per capita given by the fitted ARIMA model to compare accuracy of both methods. To compute the TFM forecasts, we need to know the data of input variable unemployment rate in the time of forecasts which we can calculate using ARIMA model methodology.

Quarterly forecasts for 2012 calculated by both methods are in Table 1. Both fitted models assume slight growth of GDP per capita in quarters of 2012. We can see that forecasted data by ARIMAX model are a little lower than simple univariate ARIMA model data.

Model	2012Q1	2012Q2	2012Q3	2012Q4
ARIMA	2.7961	3.0100	3.1961	3.1131
ARIMAX	2.6171	2.5531	2.6824	2.7212

Tab. 1 Forecasts of gross domestic product per capita by ARIMA and ARIMAX in 2012 (in thousand EUR constant prices)
(Source: Authors)

Graphical comparison of raw data of GDP per capita and fitted data and forecasts by both methodologies are at Fig. 3.

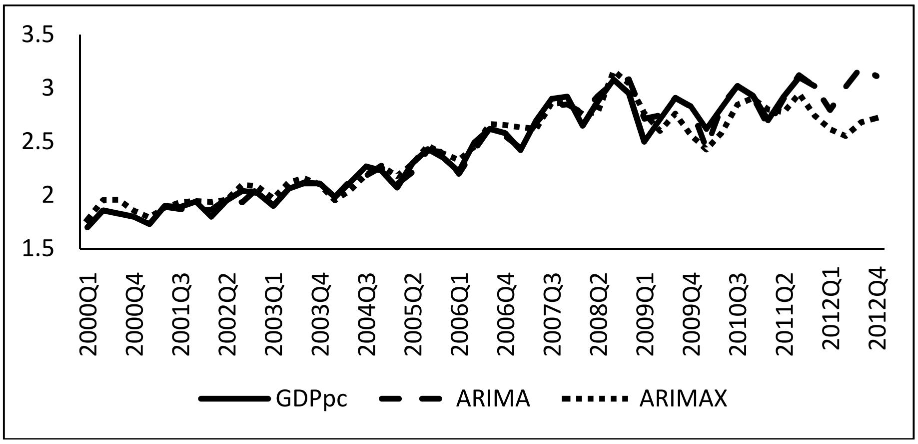

Fig. 3 Raw data of GDP per capita, ARIMA fitted data, ARIMAX fitted data and forecasts (Source: Authors)

As we can see, ARIMA model fits the trend of GDP per capita slightly better than ARIMAX model. ARIMA model mean absolute percentage error is 1.77 % and ARIMAX is 3.78 %, and root mean square error is 0.0653 respectively 0.1162.

4 Conclusions

In this article, we built the transfer function model (ARIMAX) for gross domestic product per capita as an output series and unemployment rate as an input series. Fitted model was adequate and residuals were white noise. The R-Squared statistic indicated that the model as fitted explains 92.7% of the variability in gross domestic product per capita. Next, we calculated quarterly forecasts for 2012 and compared them with forecasts calculated by simple univariate ARIMA model for GDP per capita. Forecasts were slightly different and both assumed growth of GDP per capita. ARIMA model mean absolute percentage error and root mean square error were lower than ARIMAX. ARIMA model seems to be a little accurate than ARIMAX.

Acknowledgement

This paper was elaborated in the framework of the project number IGPM 2317123: Measuring and analyzing of indicators of sustainable development using quantitative methods.

References

[1] Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.: *Time Series Analysis: Forecasting and Control*. John Wiley & Sons Inc., New York, 2008.

[2] Box, G. E. P. – Tiao, G. C.: Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association.* 1975, vol. 70, no. 349, p. 70-79.

[3] Marek, L.: Transfer function models. *Acta Oeconomica Pragnesia*. 2000, vol. 8, no. 3, p. 83-94

[4] Pankratz, A.: *Forecasting with Dynamic Regression Models*. Wiley-Interscience, 1991.

[5] Rublíková, E. – Marek, L.: Linear transfer function model for outflow rates. *Ekonomické rozhľady.* 2001, vol. 30, no. 4, p. 457-466