Prognosis of the monetary base of Mexico. Algorithm for the box-jenkins methodology (1985-2015)

ROSAS-ROJAS, Eduardo*†, ARIAS-GUZMÁN, Ericka Judith, GÁMEZ-ARROYO, Jessica, LAPA-GUZMÁN, Javier and GAVIÑO-ORTIZ, Gabriela

Universidad Autónoma del Estado de México, Km. 11.5 Carretera Atizapán de Zaragoza-Nicolás Romero S/N. Boulevard Universitario S/N Predio San Javier Atizapán de Zaragoza, Estado de México.

Received July 7, 2017; Accepted December 15, 2017

Abstract

The present research aims to provide a relevant algorithm for the proper identification of a data generating process (economic and financial variables) by using the "auto.arima" command, belonging to the forecast library (R statistical software), to identify the optimal parameters of the ARIMA model. Considering the methodological framework of Box and Jenkins process (1970), it seeks to achieve, by solving a stochastic difference equation, the correct calibration and obtaining an optimal forecast for the Mexican Monetary Base, one of the main economic variabes.

Forecast, Monetary Base, ARIMA Models

Citation: ROSAS-ROJAS, Eduardo, ARIAS-GUZMÁN, Ericka Judith, GÁMEZ-ARROYO, Jessica, LAPA-GUZMÁN, Javier and GAVIÑO-ORTIZ, Gabriela. Prognosis of the monetary base of Mexico. Algorithm for the box-jenkins methodology (1985-2015). RINOE Journal- Macroeconomics and monetary economy. 2017.1-1:34-53.

* Correspondence to Author (email: erosasr@uaemex.mx)
† Researcher contributing first author.

I. Introduction

The main motivation of this research work lies in the need for researchers and professionals in the economic and financial area to have an algorithm that is capable of developing adequate and reliable forecasts on univariate time series. In practice it is very common to have a sample of a few dozen observations that need to be predicted, even when the projection of a small number of periods is required, an algorithm that allows the calibration of an optimal forecast could be adequate, through a well defined sequence, based on the Box-Jenkins methodology and that also solves its main difficulty, that is, the correct identification of the scheme that represents the time series. Identifying not only the possible stochastic autoregressive processes and moving averages, but their possible ordinary and seasonal differences. Under these circumstances, an algorithm for forecasting becomes an essential tool.

In recent years, different statistical softwares have been developed and perfected that seek to automate, and in this way facilitate procedures for the construction of forecasts. The use of computers, information technologies, and the large amounts of data available have led institutions and organizations to use these elements together to improve research and decision making. Mathematical models (differential and difference equations) that represent in a simplified way the behavior of the event (variable) that is analyzed have also been required.

With the above elements and the proliferation of free software R, it is sought to amalgamate all the elements involved in the modeling of a data generating process to achieve its correct analysis, identification, estimation and projection. The research work is divided into six sections.

After this introduction, in the second section a review of the literature referring to the advances in the development of algorithms capable of carrying out automated forecasting is carried out; in the third, the theory referring to the ARIMA models, the concept of stationarity and the ways of checking it are addressed; in the fourth, the stages that make up the Box-Jenkin methodology are explained in detail; later, in the fifth section an empirical application is made for the time series of the Monetary Base of Mexico; in the penultimate section the results obtained from the forecast are presented and finally the conclusions are presented.

2. Review of Literature

A very common obstacle when using ARIMA models for forecasting is that the process in order selection is considered subjective and difficult to apply. In the last three decades several attempts have been made to automate the modeling of ARIMA processes.

Hannan and Rissanen (1982) proposed a method to identify the order of an ARIMA model for a series of stationary time. In his method, innovations can be obtained by adjusting a long autoregressive model for the data, and later the likelihood of the potential models is calculated through a series of standard regressions. They establish the asymptotic properties of the procedure under very general conditions.

Liu (1989) proposed a method for the identification of seasonal ARIMA models using a filtering method and certain heuristic rules; This algorithm is used in the SCA-Expert software. Gershenfeld and Weigand (1993) indicate that the selection of ARMA models is not simple. The ARMA models in general can, after choosing the autoregressive order and moving averages, adjust by means of least squares regression to find the values of the parameters that minimize the error term.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

It is generally considered good practice to find the smallest values of the processes that provide an acceptable fit to the data. Brockwell and Davis (2009) recommend using the Akaike Information Criterion (AIC) to find the appropriate number of lags for each process.

Gómez (1998) extended the identification method of Hannan-Rissanen to include the multiplicative seasonality to the identification of the ARIMA model. Gómez and Maravall (1998) implemented this automatic identification process in the SEATS software. For a given time series, the algorithm tried to find the model with the lowest Bayesian Information Criterion (BIC). Other approaches are described by Mélard and Pasteels (2000) whose algorithm for ARIMA Univariate models also allow an intervention analysis. This is implemented in the Time Series Expert software (TSE-AX).

There are some algorithms in commercial software, although these are not documented in public domain literature. They emphasize, Forecast Pro (Goodrich, 2000) that is well known for its excellent automatic algorithms for ARIMA models that were used in the M3-forecasting competition (Makridakis and Hibon, 2000). Another algorithm for commercial use is implemented in Autobox (Reilly 2000). For their part, Ord and Lowe (1996) provide a review of some commercial software that implements automatic ARIMA forecasts. Among the most used in the economic and actuarial area are: EViews, GAUSS, Gretl, Mathematica, Matlab, Minitab, SAS, Shazam, Stata, Statgraphics, SPSS.

In recent years, the R statistical software has been developing, through the paradigm of object-oriented programming, a language and environment for computational statistics and graphics, similar to the S language originally developed by Bell laboratories.

It is an open source solution for data analysis that is supported by a large and active community of researchers around the world (Kabacoff, 2011: 5). R has the following characteristics to recommend it:

- Most commercial statistical software platforms cost thousands, if not tens of thousands of pesos. R is free! If you are a teacher or student, the benefits are obvious.
- R is a comprehensive statistical platform, offering all kinds of data analysis techniques. Almost any type of data analysis can be done in R.
- R has state-of-the-art graphic capabilities. R has the most complete and powerful feature set available.
- R can easily import data from a wide variety of sources, including text files, database management systems, statistical packages and specialized data repositories.
- R provides an unprecedented platform for the programming of new statistical methods in an easy and direct way
- R contains advanced statistical routines that are not yet available in other packages. In fact, new methods are available for download on a weekly basis.

Some of the most interesting features of R are available as optional modules that can be downloaded and installed. There are more than 10,000 user-contributed modules called libraries, which provide a wide range of new algorithms. Libraries are collections of functions in R, and compiled code in a well-defined format.

Rob Hyndman (2008) has developed the "forecast" library, which is a collection of functions to develop methods and tools to display and analyze forecasts of time series that include ARIMA modeling automatically.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

December 2017 Vol.1 No.1 34-53

It is available for the R system of statistical computing (R Development Core Team 2008) is available in the network of integral files of the R project at: https://cran.rproject.org/web/packages/forecast/forecast.pdf. With the development of these and other functions it is possible to implement an algorithm capable of identifying the data generation process (PGD), to later estimate the coefficients of the integrated autoregressive processes of moving averages, which subsequently must be diagnosed to show their goodness of adjustment, and if the model is adequate considering these hypothesis tests, the last step may be the realization of a forecast.

3. Theoretical framework of time series prognostics

The theory of difference equations underlies all time series methods. In turn, time series econometrics is responsible for the estimation of equations in differences that contain a stochastic component. The proper estimation of an equation can be used for the interpretation of economic-financial data and for hypothesis testing.

The task faced by time series macroeconometricians is to develop reasonable and parsimonious models capable of forecasting, interpreting, and empirically testing hypotheses related to the stylized facts of the economic variables. This is how a methodology was developed to decompose a series into four components: trend, seasonality, cycle and an irregular component. The trend component represents the long-term behavior of the series and can be increasing or decreasing, the seasonal variation is an observed behavior with frequency less than or equal to the year, and just as the trend can be eliminated or modeled, in turn, the cyclical component represents regular periodic movements with length greater than one year, and the irregular component is a stochastic process (González, 2011: 17).

The objective of time series econometrics is to estimate and forecast these components. A time series is a sequence of data points recorded over time. For example, share prices, exports and imports, the monetary base, interest rates, the exchange rate, prices, the employment rate, international reserves, or industrial production, among others. If the data are unique numbers, the term univariate time series is used; For the vectors that consist of several series of numbers, the term multivariate time series is used.

Time series are considered to be realizations of an unknown stochastic process called the data generation process (PGD). The objective of the time series analysis is to model this process. By doing this, a formula is derived that describes the time series and can also be used frequently for forecasts (Pfaff, 2015: 20)

The methodology generally used to develop such forecasts requires finding the trajectory of the equation that directs the stochastic process and using it to predict the subsequent data. An equation in differences expresses the value of the variable as a function of its own lagged values, time and other variables. (Enders, 2015: 3). The representation of a linear difference equation of higher order with a constant coefficient is given by:

The equation is linear because all the values of the dependent variable are elevated to the first power. The term x_t is called forced process. The form of this can be very general; x_t can be some function of time, current or lagged values of other variables and / or stochastic processes. From an adequate choice of the forced process, we can obtain a wide variety of important macroeconomic models. In the econometric analysis, the forced process will contain both deterministic and stochastic components.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

The most representative cases being the following: the constant case x_t = 0, in this case we speak of an equation in differences in its traditional form; the exponential case $x_t = b(d)^{rt}$; and the deterministic trend case $x_t = b(t)^d$ and the stochastic case$x_t = \sum_{i=0}^{\infty} B_i E_{t-i}$.

ARIMA models

The time series were mainly studied under a deterministic aspect, at least until 1927, when Yule introduced the notion of the stochastic component. According to the author, each approximation to the time series can be considered as the performance of a stochastic process. This idea launched a different number of time series methods, varying in the estimation of parameters, identification, forecast and verification method.

Box and Jenkins integrated the existing knowledge and made an advance in the area creating a coherent and versatile approach that identified the three-stage iterative cycle for the time series: identification, estimation and verification of the diagnosis. The evolution of computers returned to the integrated autoregressive models of moving average (ARIMA) more popular and applicable in many scientific fields.

It is possible to combine a process of moving averages with a linear difference equation to obtain an autoregressive model and moving averages (ARMA). Consider the equation in differences of order "p". And since the forced process $\{X_t\}$ can be an MA (q) process given by: (2)

In this way:)

Following the convention of normalizing the units, this implies that B_0 is always equal to the unit. If the characteristic roots of equation (3) are all within the unit circle, $\{Y_t\}$ is known as the ARMA model for Y_t. The autoregressive part of the model is the differential equation given by the homogeneous portion of (2) and the part of the moving average is the sequence $\{X_t\}$. If the homogeneous part of the equation in differences contains "p" lags and the model for X_t contains q lags, the model is called ARMA (p, q). If q = 0. The process is called pure autoregressive, denoted by AR (p), and if p = 0, the process is a pure moving average, denoted by MA (q).

In an ARMA model it is perfectly permissible that p and / or q are infinite. If one or more of the characteristic roots of (3) is greater than or equal to the unit, the sequence $\{Y_t\}$ is said to be an integrated process and in this case the model is called Integrated Self-Regressive Moving Averages (ARIMA) (Enders). , 2015: 50).

If we treat equation (3) as a difference equation, it can be solved for Y_t in terms of the sequence $\{E_t\}$. For the general model ARMA (p, q) of equation (3) when using the lag operator we have to: The sum will lead to an MA (∞) process. The important thing is to determine if such expansion is convergent with what the stochastic difference equation given in (5) is stable[6].

Stationery

In formal terms, the stationarity of a time series is defined in two ways. The first concept of stationarity is in a strong or strict sense.

[6] The stability condition is that the roots of the polynomial $\left(1 - \sum_{i=1}^{p} a_i L^i\right)$ they must fall outside the unit circle. This would also show that Y_t is a linear stochastic difference equation, so the stability condition is necessary for the time series $\{Y_t\}$ be stationary.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

RINOE Journal
December 2017 Vol.1 No.1 34-53

It is said that a time series Y_t is stationary in a strict or strong sense, if its function of joint probability distribution remains identical in time. That is, it is stationary in the strict sense, if and only if:

$$f(Y_t, Y_{t-1}, \dots, Y_{t-k}) = f(Y_{t+T}, Y_{t-1+T}, \dots, Y_{t-k+T}), \ \forall T \in \mathbb{R} \quad (1)$$

While stationarity in a weak or broad sense. It is said that a series of time is stationary in a broad or weak sense, if and only if it satisfies three conditions:

1. Your average is constant in time. $E(Y_t) = E(Y_{t+T}) = \mu \ \ \forall T \in \mathbb{R}$. If there is no trend, it can be assumed that the mean is constant and that the observed value for each period can be represented by a constant that will be the sample mean.
2. The variance is constant over time. The dispersion of the process is also invariant in time. $Var(Y_t) = Var(Y_{t+T}) = \gamma \ \ \forall T \in \mathbb{R}$.
3. The autocorrelation function is independent of time. This measures the possible dependence between an observed value (Y_t) and another (Y_{t-k}) that is separated by a range of length k.

If the time series is stationary, then the values of the mean, variance and autocorrelation function can be estimated by considering the time series as a sample of defined size, whose parameters are constant over time. Otherwise, when the time series is non-stationary, the problem is that the different samples represent samples of size one, which is not representative, so it is practically impossible to infer the properties of the population from these data (González, 2011: 53).

Stationarity tests

There are several tests to identify if a certain time series is stationary, among the most used in a prominent way in the literature are: 1) Graphic analysis; 2) The correlogram test, and its corresponding simple autocorrelation function (FAC) and partial autocorrelation function (FACP); and 3) The unit root test. The last two are explained as the fundamental basis of the analysis of ARIMA models.

Simple and Partial Autocorrelation Function

Autocovariances of each time point divided by the initial autocovariance ($p_s = \frac{\gamma_s}{\gamma_0}$) produces the autocorrelations (table 1), which in turn serve as a mathematical and visual representation (correlogram) in the Box-Jenkins methodology to identify and estimate the time series models. The autocorrelation function (FAC) should depend only on the "s" interval between two variables and not on its position at time t. For its part, the Partial Autocorrelation function (FACP) measures the marginal contribution or the weight of each new autoregressive variable that is added to the model. Table 1 shows some of the most representative stochastic processes.

Model	Representation	Equation
White noise	ARIMA(0,0,0)	$\Delta Y_t = E_t$
Random Walk without Drift	ARIMA(0,1,0)	$Y_t = Y_{t-1} + E_t$
Random Walk with Drift	ARIMA(0,1,0)	$Y_t = a_0 + Y_{t-1} + E_t$
Autoregressive	ARIMA(p,0,0)	$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + E_t$
SMA	ARIMA(0,0,q)	$Y_t = a_0 + E_t + B_1 E_{t-1} + \cdots + B_q E_{t-q}$
Autoregressive Integrated Mobile Media	ARIMA(1,1,1)	$\Delta Y_t = a_0 + a_1 \Delta Y_{t-1} + E_t + B_1 \Delta E_{t-1}$

Table 1 Representation of various Stochastic Processes
Source: self made

For each of the different stochastic processes, the corresponding mean, variance and autocovariance must be obtained, since this will allow determining if the determined stochastic process is stationary or not.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

December 2017 Vol.1 No.1 34-53

By dividing the autocovariance between the variance, the corresponding simple autocorrelation function is obtained, which is independent of time. Below is a Table (2) summary of the first two statistical moments and the autocovariance of the most common stochastic processes.

Proceso	Media	Varianza	Autocovarianza
AR(1)	$\frac{a_0}{1-a_1} = u$	$\frac{\sigma^2}{1-a_1^2} = \gamma_0$	$\frac{\sigma^2 (a_1)^S}{1-a_1^2} = \gamma_S$
AR(2)	$\frac{a_0}{1-a_1-a_2} = u$	$\frac{(1-a_2)\sigma^2}{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)} = \gamma_0$	$a_1 \gamma_{s-1} + a_2 \gamma_{s-2} = \gamma_S$
AR(p)	$\frac{a_0}{1-a_1-a_2-\cdots-a_p} = u$	$a_1\gamma_1 + a_2\gamma_2 + \cdots + a_P\gamma_P + \sigma^2 = \gamma_0$	$a_1\gamma_{S-1} + a_2\gamma_{S-2} + \cdots + a_P\gamma_{S-P}$
MA(1)	$a_0 = u$	$(1+B_1^2)\sigma^2 = \gamma_0$	$\gamma_1 = B_1\sigma^2, \quad y\, \gamma_S = 0, \forall$
MA(2)	$a_0 = u$	$(1+B_1^2+B_2^2)\sigma^2 = \gamma_0$	$\gamma_1 = (B_1 + B_1B_2)\sigma^2;$ $\gamma_2 = (B_2)\sigma^2;$ $\gamma_S = 0, \forall\, s > 2$
MA(q)	$a_0 = u$	$(1+B_1^2+B_2^2+\cdots+B_q^2)\sigma^2 = \gamma_0$	$B_S\sigma^2 = \gamma_S$
ARMA(1,1)	$a_0 = u$	$\frac{(1+B_1^2+2a_1B_1)\sigma^2}{1-a_1^2} = \gamma_0$	$\frac{(1+a_1B_1)(a_1+B_1)\sigma^2}{1-a_1^2}$

Table 2 Mean, variance and autocovariance of different processes
Source: self made

For example, in an AR (1) process, Y_t and Y_{t-2} are correlated even when Y_{t-2} does not appear directly in the model. The correlation between Y_t and Y_{t-2}, ie p_2, is equal to the correlation between Y_t and Y_{t-1}, ie p_1 multiplied by the correlation between Y_{t-1} and Y_{t-2}, p_1 again, so $p_2 = (p_1)^2$. It is important to note that such indirect correlations are present in the Simple Autocorrelation function of any autoregressive process.

On the other hand, the partial correlation between Y_t and Y_{t-S} eliminates the effect of the values that intervenes from Y_{t-1} up to and Y_{t-S+1}. In such a way that, in an AR (1) process, the Partial Autocorrelation between Y_t and Y_{t-2} equals zero.

Raíz Unitaria (Prueba Dickey Fuller Aumentada)

Assuming an ARIMA autoregressive model (1,0,0) whose algebraic representation is: $Y_t = a_1 Y_{t-1} + e_t$, subtracting Y_{t-1} from each side of the equation to write the process in an equivalent way like: $\Delta Y_t = \delta Y_{t-1} + e_t$, where $\delta = a_1 - 1$. Test the hypothesis that $a_1 = 1$, is equivalent to testing the hypothesis that $\delta = 0$. Dickey and Fuller (1979), considered three different regression equations that can be used to prove the presence of unit root.

The difference between the three regressions has to do with the presence of the deterministic elements a_0 and a_2 t. The first is a pure random walk model, the second includes an intercept or a drift term, and the third includes both a drift and a linear temporal trend. The parameter of interest in all the regression equations is δ; if $\delta = 0$, sequence $\{Y_t\}$ contains a unit root The test involves estimating one or more of the equations described above using ordinary least squares (OLS) to obtain the estimated values of δ and its associated standard error. The result obtained is a distribution function called "tau" that allows determining if the null hypothesis should be rejected or not.

Dickey and Fuller (1981), developed a test in which the error term was correlated, and was called Dickey Fuller Augmented Test (DFA). This test involves increasing the three previous equations with the lagged values of the dependent variable ΔY_t in such a way that the processes are determined as follows, Dickey and Fuller (1981), developed a test in which the error term was correlated, and was called Dickey Fuller Augmented Test (DFA). This test involves increasing the three previous equations with the lagged values of the dependent variable in such a way that the processes are determined as follows:

ISSN-On line: 2524-2040
RINOE® All rights reserved.

The process e_t It is a white noise. The term of lagged differences that must be included is identified by information criteria of Akaike and Schwarz. In these new regression models, the idea remains the same, that is, to prove that $\delta = 0$, based on the same tau distribution [7].

4. Methodology Box-Jenkins. Prognostics

The Box-Jenkins methodology for modeling the processes (Integrated Self-Regressive of Moving Averages) ARIMA was described in a highly influential book by statisticians George Box and Gwilym Jenkins in 1970. An ARIMA process is a mathematical model used to forecast. The Box-Jenkins methodology involves identifying an appropriate ARIMA process, adapting it to the data and then using the adjusted model for prediction. One of the attractive features of the Box-Jenkins approach to prediction is that ARIMA processes are a very rich class of possible models and it is usually plausible to find a process that provides an adequate description of the data (Hyndman, 2008).

The original Box-Jenkins methodology involves an iterative process of stages corresponding to the selection of models, parameter estimation and verification of the same. Recent explanations of the process often add a preliminary stage of data preparation and a final stage of application of the model (or forecast) (Makridakis, Wheelwright and Hyndman, 1998). The Box-Jenkins approach to construct linear models of time series is based on two main principles: The first, the principle of parsimony, which consists of always choosing the simplest model that is sufficiently representative of the data; and the second, the principle of iterative improvement, which consists of starting from a simple and feasible model, to which successive improvements will be made, until arriving at a satisfactory model.

The Box-Jenkins methodology consists of the following stages:

Analysis of Stationarity and Seasonality

In the first stage it must be verified that the series is stationary and try to eliminate the seasonal variation, if it is the case. González (2011: 85) suggests the following order to deal with non-stationary models:

a. Stabilization of variance. To make the data homoscedastic, mathematical transformations are suggested such as: The square root, the neperian logarithm, the reciprocal or inverse, the reciprocal of the square root and the reciprocal of the logarithm. Another common transformation in economic series is to deflate the data.

b. Elimination of the trend. We proceed to review if the data present some kind of trend. If it exists, it will be eliminated through ordinary differences.

c. Treatment of seasonal variation. In some data it happens that the seasonal variation is visible from the graph of the original data or by stabilizing the variance. In other data it is possible to be warned until after the trend has been eliminated. If the seasonal variation is detected, it must be eliminated with seasonal differences or it must be represented by multiplicative models that include the seasonal component SARIMA (p, d, q) x $(P, D, Q)_m$, with "M" denoting seasonality.

[7] The levels of significance without constant at 1%, 5% and 10% are: -2.58, -1.95 and -1.62, respectively; with constant: -3.43, -2.86 and -2.57, while with constant and trend: -3.96, -3.41 and -3.12.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

d. Where m is the number of periods per station, capitalization notation is used to denote the seasonal part of the model and lowercase notation for the non-seasonal part of the model. The seasonal part of the model consists of terms that are very similar to the non-seasonal component, but these imply seasonal lags of the period. For example, a model SARIMA (1,1,1) x $(1,1,1)_{12}$, without constant for monthly data can be written as follows:

Identification of the Tentative Model

The correlograms are used for the graphical and statistical representation of the FAC and FACP, with the purpose of determining a pattern of the different autocorrelations[8], It will also be verified by hypothesis tests such as the Q statistic developed by Box and Pierce (1970) or the Ljung-Box statistic (1978). In the same way, the correlogram is used to evaluate if there is some kind of seasonal variation, not eliminated by seasonal differences.

Estimation of Model Parameters

For each tentative model the parameters must be estimated. This can be done using the ordinary least squares method, the Yule-Walker equation or the maximum likelihood method. The estimated values must be examined for significance and stability. If the model is not stable (implying an explosive trajectory) or the coefficients do not differ significantly from zero, the model should be re-examined and a new estimation of it should be proposed.

Verification of the Model

It is recommended to carry out the following analysis:

a. Classic statistics must be calculated to measure the fit of the data, this allows to compare the different models, both in their adjustment and in their forecasting capacity. Also known as error measures[9].

b. Perform the individual hypothesis tests (t-statistics) on the statistical significance of the estimated coefficients.

c. Carry out the tests on the residuals of the model, and obtain their simple and partial correlograms, to identify that they are stationary. Additionally, perform an Augmented Dickey-Fuller test (1979) to determine the stationarity of the disturbances.

d. Prove that there is no autocorrelation, for this purpose the test called "portmanteau" by Box-Pierce, and / or Ljung-Box is used.

Forecast

The most important use of an ARIMA model is to predict the future values of the sequence $\{Y_t\}$. The forecasts will be constructed using the advanced iteration method. It is important to note that the quality of the forecast declines as it continues.

[8] The statistical significance of any coefficient of autocorrelation is judged by its standard error. Bartlett showed that if a time series is purely random, that is, if it is a sample of white noise, the sample autocorrelation coefficients are approximately: $\rho k \sim N(0, 1/n)$.

Therefore, according to the properties of the standard normal distribution, the confidence interval of 95% for any (population) ρk es: $\rho k \pm 1.96(\sqrt{(1/n)})$.

[9] ME= (Mean Error) Error Medio; RMSE= (Root Mean Squared Error) Raíz Cuadrada del Error Cuadrático medio; MAE=(Mean Absolute Error) Error Absoluto Medio; MPE= (Mean Percentage Error) Error Medio Porcentual; MAPE=(Mean Absolute Percentage Error) Error Absoluto Medio Porcentual.

RINOE® All rights reserved.

December 2017 Vol.1 No.1 34-53

Validation of the forecast

One way of knowing which of the selected models has the best performance to forecast is to develop a range of summary measures on the accuracy of the forecast based on the error measures. Ideally, it is desirable that forecast errors have an average close to zero and minimum variance. A Diebold-Mariano test (1995) must be carried out, in this test an objective function that is not quadratic is sought.

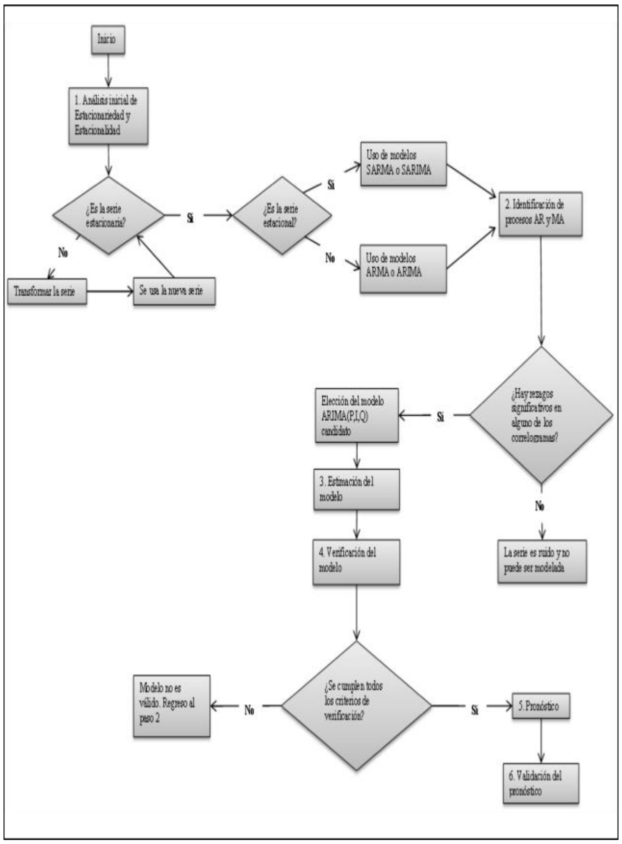


Figure 1 Flowchart of the Box-Jenkins Methodology
Source: self made

5. Empirical Application for the Monetary Basis of Mexico

Next, we apply an algorithm based on the Box-Jenkinsmethodology to identify the behavior of the time series of the Mexican Monetary Base. With the help of several libraries belonging to the R system of statistical computing, the aim is to design a procedure that allows obtaining the optimal model, for which we will use one of the most important real variables of any economy, such as the Monetary Base.

Since the reform of the Bretton Woods agreement in 1976, it was established that exchange rate regimes would be determined flexibly by supply and demand. In this way, the creation of international liquidity remained in the hands of international financial markets. This agreement has generated international liquidity faster than the growth of the world economy and international trade (Mántey, 2013: 63). The disproportionate increase in the currency induces mainly speculative capital flows that perniciously burst from one economy to another, and mainly affecting the emerging economies.

International short-term capital flows cause volatility in exchange rates, in which emerging economies through their central banks must intervene in their foreign exchange markets in order to avoid strong appreciation of their currency and subsequently sterilize such interventions with which avoids inflationary pressures. Sterilization is the process by which the monetary authority ensures that interventions of the exchange rate will not affect the domestic monetary base, which is one of the components of the money supply in general. The accumulation of international assets has become the main support of the monetary base in the Mexican economy, while the share of domestic credit that the central bank grants to the government and commercial banks has been reduced.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

In the simplified balance sheet of any monetary authority, the monetary base is the sum of the credit of the central bank (government and commercial banks) plus its international reserves. In such a way that an increase in any of them will cause an increase in the money supply and vice versa. In a country like Mexico, determining the optimal economic policy in an environment of considerable income from speculative capital is a difficult task. Opting for a policy of intervention in the exchange market generates an increase in the currency of local currency, which could lead to a higher level of inflation, however, the Bank of Mexico can reduce these inflationary pressures by sterilizing interventions.

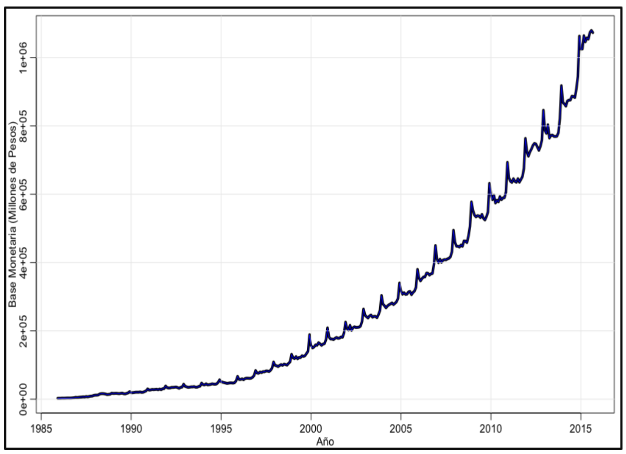
If inflation is high, sterilization of interventions may be necessary to meet a certain inflation target and not to affect the credibility of monetary policy (Borensztein, 2015: 13). On the other hand, Hüfner (2004) points out that, in developing countries, sterilized interventions can contribute to a better control of inflation than would be possible using the interest rate as the sole instrument of monetary policy. Any nation that adopts a fixed exchange rate regime or managed float, and also performs a continuous accumulation of international reserves, resulting in an increase in the monetary base, must inevitably rely on the exchange interventions made by its monetary authority as a means to maintain the stability of the exchange rate, and thereby achieve your inflation goal.

Description of Data

The monetary base consists of the bills and coins in circulation and the total net balance of the current accounts that the credit institutions maintain in the Bank of Mexico. In our country the behavior of the monetary base is very similar to that of banknotes and coins in circulation because, as a rule, banks do not try to maintain positive balances in their accounts in the issuing institute (Banxico, 1998: 4).

The Bank of Mexico satisfies daily fluctuations in the demand for banknotes from the public. To do this, it has to create (buy currencies, securities and grant credit to commercial banks) or destroy (sell currencies, securities and become a debtor to the bank) monetary base, as the case may be. All the changes in the supply and demand of Monetary Base, are necessarily reflected in variations in the balances of the current accounts of the bank in the Bank of Mexico. To avoid pressure, the central institute balances the supply daily with the demand for monetary base through its daily interventions in the money market. This is done every working day from 12:00 hours (Banxico, 1998: 8).

The database from which the time series is compiled is: International Financial Statistics (IFS) of the International Monetary Fund. The periodicity is monthly and includes the period from December 1985 to September 2015, the units of the 358 data are given in millions of pesos, with this it is sought to have a considerable length in the database, which allows obtaining more estimates precise.



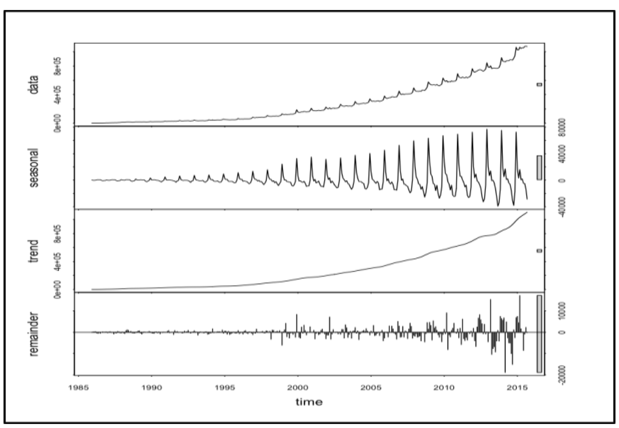
Grph 1 Monetary Base of Mexico
Source: Prepared by the author with data from the International Financial Statistics of the IMF

ISSN-On line: 2524-2040
RINOE® All rights reserved.

The monetary base (graph 1) presents a very defined seasonal pattern, through which it increases drastically towards the end of the year, when bonuses are paid and sales increase during the Christmas season. At the beginning of the cycle, demand decreases immediately and then suffers smaller variations. Even, it increases depending on the payment of the fortnights of the companies, the payment of taxes, the entrance of vacations and other effects of the calendar (Heath, 2012: 318).

Estimation of the Box-Jenkins methodology

According to the Box-Jenkins methodology, in the first stage of the algorithm the Stationarity and Seasonality Analysis must be performed. Using the "stl" command we decompose the series into its components: trend, seasonality and noise (graph 2). What is observed is that the time series of the monetary base has a strong seasonal component, a clear positive trend and a remnant whose variation has increased since 1999. Likewise, the simple and partial correlograms were calculated and showed that the series behaves as non-stationary.



Graph 2 Components of the time series of the Monetary Base
Source: Own elaboration with R 3.3.3

The Augmented Dickey Fuller test is also carried out in the algorithm to identify whether the time series has a unit root. Using the "urdfTest" command from the library "fUnitRoots" (Wuertz, 2009), it is verified that the Monetary Base is a non-stationary series. The "forecast" library developed by Hyndman (2008) provides the commands "ndiffs" and "nsdiffs" that allow estimating the number of differences required for a series of time to become stationary, in the same way the number of differences can be estimated seasonal The results of both tests indicate that the series must be differentiated ordinarily and seasonally, in both cases the order is one.

Considering the previous analysis, we proceed to stabilize the variance applying Neperian logarithms, the trend is eliminated by ordinary differentiation and the model is represented with a seasonal component. The results of the Augmented Dickey Fuller test are presented in Table 3.

Variable	**Condicionantes**	**Tau Statistic**	**p-value**
Monetary base	Without Intercept	4,5321	NA
	Intercept	2,8239	NA
	Intercept and Trend	-0,1049	0,9165
Δlog(Monetary base)	Without Intercept	-13,8685	0,0000
	Intercept	-15,5037	0,0000
	Intercept and Trend	-15,9514	0,0000

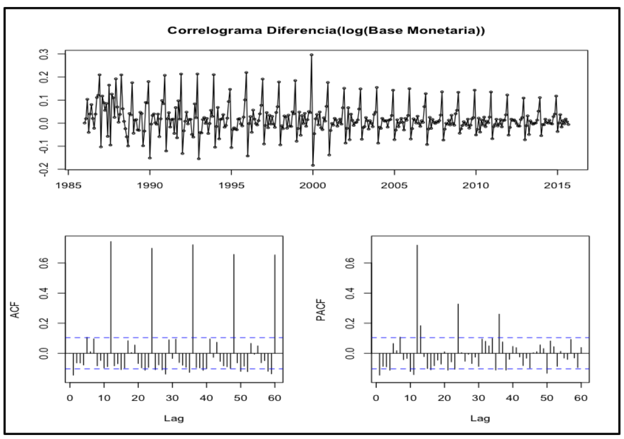
Table 3 Augmented Dickey-Fuller test to identify Unitary Root
Source: Own program elaboration R 3.3.3

In the second stage of the algorithm, the simple and partial correlogram of the variable transformed with logarithmic differences are used, this allows to identify the pattern that represents the stochastic process, and in this way, to propose the appropriate ARIMA model. There is also a marked seasonal variation, which is present in the twelfth month of each year.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

As already mentioned, the "forecast" library has the function "auto.arima" that allows defining the parameters p, d, q considering different information criteria such as Akaike (AIC, for its acronym in English) and Schwartz Bayesiano (BIC, for its acronym in English)[10], the latter will always select a more parsimonious model than the first one (Enders, 2015: 69).

For the time series of the Monetary Base of Mexico the function indicates that it is a model $SARIMA(0,1,2)x(0,1,3)_{12}$. This means that the stochastic process is represented by an order of integration one, a component of ordinary moving averages of order two and another component of seasonal integration, in addition to a component of seasonal moving averages of order three. The above corroborates what was established in the first stage, with respect to the strong presence of the seasonal component (graph 3).



Grph 3 Function of Simple and Partial Autocorrelation of the Monetary Base
Source: Own program elaboration R 3.3.3

In the third stage of the algorithm, the estimated coefficients of the proposed model are presented, and two other possible models are presented to confirm that the stochastic process indicated by the "auto.arima" function effectively provides the optimal model for the time series. The method programmed in the algorithm is Maximum Likelihood.

	Dependent variable: Δlog(Monetary base)		
	Model(1)	**Model(2)**	**Model(3)**
	SARIMA (0,1,2)x(0,1,3)[12]	SARIMA (0,1,2)x(0,1,2)[12]	SARIMA (0,1,2)x(0,1,4)[12]
MA(1)	-0.2728***	-0.2870***	-0.2752***
	-0,0555	-0,0552	-0,0556
MA(2)	0.1616***	0.1971***	0.1638***
	-0,0566	-0,0552	-0,0564
SMA(1)	-0.6031***	-0.6003***	-0.6136***
	-0,0647	-0,0625	-0,0675
SMA(2)	-0,1325	0,1702	-0,1263
	-0,0858	-0,052	-0,0931
SMA(3)	0,3419		0.3231***
	(0.0810)***		-0,0914
SMA(4)			0,0304
			-0,0584
Observations	358	358	358
AIC	7077,98	7088,99	7079,71
BIC	7101,05	7108,21	7106,62
log likelihood	-3532,99	-3539,49	-3532,86
ME	779,739	803,3763	778,3933
RMSE	6561,751	6737,466	6554,812
MAE	3980,888	4175,795	3968,208
MPE	0,2931	0,2877986	0,2931

Table 4 Results of the SARIMA Regression (Maximum Likelihood)
Source: Own program elaboration R 3.3.3

Table 4 presents the results of the coefficients for the different proposals for stochastic processes. The model (1) is the one proposed as optimal, while model (2) has been constructed without considering the term of the seasonal moving average for month 36, and for the model (3).

[10] $AIC = T * \log(SRC) + 2 * n$
$BIC = T * \log(SRC) + n * \log(T)$
Where n= number of estimated parameters (p + q + constant term) and T= number of observations used.

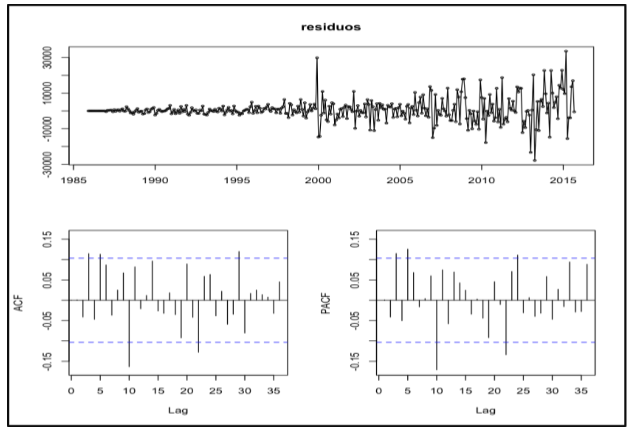
ISSN-On line: 2524-2040
RINOE® All rights reserved.

December 2017 Vol.1 No.1 34-53

The same coefficients of the optimal model have been used but a seasonal moving average has been added that captures the value of the estimator corresponding to month 48. According to the Box-Jenkins methodology, in the next stage an analysis must be carried out to diagnose the goodness of fit of the optimal model, in the previous table the main statistics used in practice are recorded, based on the sum of quadratic residuals (SRC). It is evident that models (1) and (3) present the lowest values of these statistics (ME, RMSE, MAE, MPE), and that most of the estimated coefficients are in both cases statistically significant.

However, the information criteria of Akaike and Bayesiano indicate that the model that best estimates the true parameters is the model (1). We can identify that the two ordinary moving averages are statistically significant at a confidence level of 99%, while two of the three seasonal moving averages were statistically significant with very high confidence levels. It is also important to point out that the model does not present an intersection coefficient, this is because when this coefficient was included it was not statistically significant.

Additionally, since the model is explained by stationary stochastic processes, it is guaranteed to be stable. With the previous empirical evidence we proceed to graph the residuals of the selected model, its corresponding autocorrelation functions guarantee that it is a stationary process (Graph 4). The previous thing is also corroborated with the unit root test of Dickey and Fuller in its three versions (-13.52, -13.79 and -14.11, without intercept, with intercept and with tendency and intercept, respectively), which indicates that the residuals of the optimal model are randomly distributed.



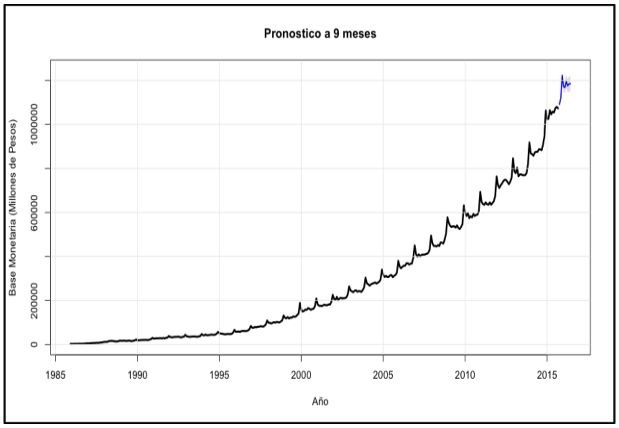
Graph 4 FAC and FACP of the residues of the optimal model
Source: Own program elaboration R 3.3.3

To ensure that the residues of the optimal model behave independently, statistical algorithms were programmed in the algorithm to guarantee non-autocorrelation, the hypotheses were developed based on Box-Pierce and Ljung Box, both show that there is no autocorrelation in the first 5 lags Once all the criteria of the diagnosis have been verified, it is possible that the algorithm makes the forecast of the future values of the sequence that follows the Monetary Base.

6. Results

Once the compliance with the assumptions about the residuals of the optimal model and the other proposed models has been verified, they are calibrated to make forecasts for the next 9 months of the time series of the Mexican Monetary Base.

ISSN-On line: 2524-2040
RINOE® All rights reserved.



Graph 5 Forecast for 9 months of the Monetary Base
Source: Prepared by the author with the R program 3.3.3

One way to identify which of the selected models has the best performance to forecast is to develop a range of summary measures on the accuracy of the forecast based on the error measures. Ideally, it is desirable that forecast errors have an average close to zero and minimum variance. However, if the difference in constructed forecasts is statistically significant, then it can not be efficiently diagnosed by the aforementioned measures.

To solve this problem the algorithm includes a modern method of statistical evaluation, known as the Diebold-Mariano test[11] (1995), that offers a quantitative evaluation tool of the accuracy of the forecast for the Monetary Base of Mexico.

	DM test based on Model (2) and Model (1)	DM test based on Model (2) and Model (3)	DM test based on Model (1) and Model (3)
D-M statistics	1,6851	1,8699	1,2341
p-values	0,04642	0,03116	0,109

Table 5 Diebold-Mariano test

Note: The statistical test is based on Quadratic Absolute Errors, whose objective function is quadratic. Source: Prepared by the author with the program R 3.3.3

From table 5 the conclusions of the comparison of the models (1), (2) and (3) establish that according to the Diebold-Mariano test the statistic is 1.6851, which rejects the null hypothesis at the level of significance 5%, that is, the differences observed between the forecasts developed is significant and the accuracy of the forecast of the model (1) is better with respect to the model (2).

The same occurs with the test developed between the model (3) and the model (2), whose statistic is 1.8699, indicating that the model (3) provides a better fit than the model (2). Finally, the Diebold-Mariano test between models (1) and (3) shows that the statistic is 1.2341, so the null hypothesis can not be rejected at the 10% level of significance, that is, the difference observed between the Prognostics of both models is not statistically significant, noting that they have the same accuracy in the forecast. However, given that the AIC and the BIC (this criterion provides the most parsimonious model) they indicate that the optimal option should be the model (1), that is why it is used for the calculation of the forecast.

Then, table 6 shows the projection by interval of the 9 months after the last observed value, the lower and upper limits are calculated (Minimum and maximum value, respectively) of a range constructed with a confidence level of 95%, to calculate it, the point estimator (BM Forecast) was used as a center.

[11] The null hypothesis is that the two methods have the same accuracy in the forecast, while the alternative hypothesis is that method 2 is more accurate than method 1. (In R-Project: Alternative hypothesis = "greater").

RINOE® All rights reserved.

Article

RINOE Journal

December 2017 Vol.1 No.1 34-53

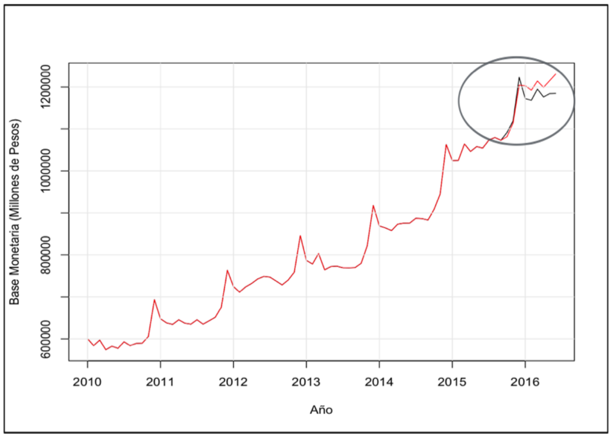
In addition, the true value recorded by the Mexican Monetary Base for the period of October 2015 to June 2016, provided by the Bank of Mexico, is presented. [12].

Mes/Año	Minimum value	BM Forecast	Maximum value	Real Values
October 2015	1078469	1091666	1104863	1081587,28
November 2015	1103176	1119493	1135810	1114508,08
December 2015	1202934	1223029	1243124	1204344,70
January 2016	1149077	1172345	1195613	1203142,75
February 2016	1142278	1168334	1194391	1192508,08
March 2016	1166495	1195070	1223644	1214379,10
April 2016	1145245	1176133	1207021	1199118,36
May 2016	1150987	1184027	1217067	1215015,55
June 2016	1149587	1184646	1219706	1230752,88

Table 6 Forecast Monetary Base of Mexico (Millions of Pesos)
Source: Own elaboration with figures predicted with the R program 3.3.3. the Real Securities are taken from Banxico.

According to the real values presented by Banxico, the prediction of the model (1) is quite accurate, the first three months are predicted effectively, while in the fourth month the prospection is outside the maximum limit for a level of confidence 95%, this is explained because, as expected, the forecasts are more accurate in the short than in the long term. However, the following 4 months, from February to May 2016, are also predicted within the established limits of significance, and finally the ninth month is again outside the maximum limit.

In general, we can establish that the algorithm works properly with regard to the optimal identification of the model, its calibration to estimate the coefficients of the different stochastic processes and the construction and verification of the forecast.



Graph 6 Real Values vs. Predicted Values
Source: Own program elaboration R 3.3.3

Graph 6 shows the visual comparison between the real values and the predicted values, evidencing that there is a very good adjustment in the first months, practically the values overlap, and that later, despite the bias in the forecast of the fourth month the slopes positive and negative, which actually happened, were also correctly identified by the model built through the algorithm.

7. Conclusions

The present research shows that the proposed algorithm is effective in obtaining an accurate and reliable forecast for the first three months. The described procedure can be applied to different economic and financial variables to obtain reliable forecasts and thereby support decision making.

The forecasts obtained using the Box-Jenkins methodology have proven to be more reliable than those obtained from traditional econometric models, particularly in the case of short-term projections, as observed in the empirical application of this work.

[12] The data can be consulted at the following electronic address: http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?accion=consultarCuadro&idCuadro=CF104&sector=3&locale=es

ISSN-On line: 2524-2040
RINOE® All rights reserved.

The adequate analysis, the correct identification, the calibration of the estimate, together with a clear validation, allow the optimal construction and verification of the forecast of a variable of vital importance for the Mexican economy, such as the monetary base, of the predicted behavior You can identify behaviors regarding the way in which the exchange and monetary policies in the Mexican economy are applied. Therefore, the subsequent work in this materia, will require adding to the forecast the exogenous variables that capture these behaviors.

A large proportion of the monetary base is offered during the intervention process in the currency market by the Bank of Mexico. To achieve the objectives of its monetary policy and alleviate inflationary pressures, the Central Bank intervenes in the foreign exchange market and at the same time actively adopts sterilization measures to neutralize the impacts of excess liquidity due to the increase in foreign currency reserves through changes Opposite in the net internal assets (internal credit of the Central Bank to the Government), this is evidenced by the clear pattern of constant growth of the money supply that has been presented and that is predicted to continue in the future. It is also clear that the sterilized intervention policy remains the tool through which the Central Bank achieves its goals of inflation targeting.

The constant growth of the monetary base shows that Banxico effectively sterilizes the impacts of speculative capital inflows and increases foreign exchange reserves in the national money supply. That is, the large-scale intervention of the foreign exchange market and the sterilization operations have a significant influence on Mexico's money supply.

In this context, it is evident that the current and predicted increase in the monetary base is aimed at maintaining the solvency of the banking system, the stability of the interest rate and the appreciation of the exchange rate to maintain internal and external balance. Consequently, the demand for the monetary base is determined by the needs for balance compensation between banks and the need for precautionary monetary balances, determined by the latter, by the bankers' state of trust (Mantey, 2004).

Aknowledgement

The research was developed within the framework of the project: "Analysis of the impact of the exchange policy on the Mexican economy" 5658 / UAEM-CA-96.

References

Banco de México (1998). La conducción de la Política Monetaria del banco de México a través del Regimen de Saldos Acumulados, Dirección General de Operaciones de Banca Central.

Borensztein, E., Hernaiz, D.,Pineda,E, y A. Rasteletti,(2015). Intervención Cambiaria y Costos de Esterilización: Algunas Reflexiones de Política. *Resumen de políticas del BID,* (241).

Box, G. E. P. & D., A. Pierce, (1970), Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65, 1509–1526.

Box, G.E.P. & G.M. Jenkins., (1970), *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.

Brockwell, P.J. and R.A. Davis (2009). Time Series: Theory and Methods, 2nd ed. Springer.
Dickey D. A., y W.A. Fuller, (1979).

ISSN-On line: 2524-2040
RINOE® All rights reserved.

"Distribution of the Estimators for Autoregressive Time Series with a Unit Root", en Journal of the American Statistical Association, vol. 74, pp. 427-431.

Dickey, D. A., and W. A. Fuller. (1981). "Likelihood Ratio Tests forAutoregressive Time Series with a Unit Root." Econometrica, 49, pp. 1057–1072.

Diebold, F.X. & R. Mariano (1995). Comparing predictive accuracy. *J. Bus. Econ. Stat.*, *13*, 253–265.

Enders, W. (2015). Applied Econometric Time Series. New York, NY:Wiley.

Gershenfeld, N.A. and A.S. Weigand (1993). "The Future of Time Series", Learning and Understanding. Time Series Prediction. Forecasting the Future and Understanding the Past. In: Eds. A.S.Wigand and Gersehenfeld, N.A. SFl Studies in the Sciences Complexity, vol. IV, Addison Wesley, pp. 1-70.

Gómez V (1998). "Automatic Model Identi_cation in the Presence of Missing Observations and Outliers." Working paper D-98009, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.

Gómez V & A. Maravall. (1998). "Programs TRAMO and SEATS, Instructions for the Users". Working paper 97001, Ministerio de Economía y Hacienda, Dirección General de Análisis y Programación Presupuestaria.

González V., M. (2011). "Pronósticos: metodología Box-Jenkins", UNAM. México.

Goodrich R. L. (2000). The Forecast Pro Methodology." International Journal of Forecasting, 16(4), 533-535.

Hannan E.J., Rissanen J. (1982). "Recursive Estimation of Mixed Autoregressive-Moving Average Order". Biometrika, 69(1), 81-94.

Heath, J. (2012). Lo que indican los indicadores. como utilizar la información estadística para entender la realidad económica de México. México, D.F.: INEGI.

Hüfner, F. (2004), Foreign Exchange Intervention as a Monetary Policy Instrument: Evidence for Inflation Targeting Countries, ZEW Economic Studies 23, Centre for European Economic Rsearch, Physica-Verlag, Heidelberg.

Hyndman, R. J., & Y. Khandakar. (2008)."Automatic time series forecasting: the forecast package for r". Journal of Statistical Software, 26(3).

Kabacoff, R. (2011). R in Action. Manning Publications Co.

Liu LM. (1989). "Identication of Seasonal Arima Models Using a Filtering Method". Communications in Statistics: Theory & Methods, 18, 2279-2288.

Ljung, G. M. & G., E., P. Box,(1978). On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.

Makridakis S, Hibon M (2000). "The M3-Competition: Results, Conclusions and Implications". International Journal of Forecasting, 16, 451-476.

Makridakis, S., & S.C. Wheelwright, and R.J. Hyndman., (1998). *Forecasting: methods and applications*, New York: John Wiley & Sons.

Mántey G. 2004, "La política monetaria en México"; en Eugenia C., Alicia G. y Alma CH. (comps), *Economía financiera contemporánea*, Miguel Ángel Porrúa-UNAM, México.

ISSN-On line: 2524-2040
RINOE® All rights reserved.

Mántey, G. (2013). “El apoyo fiscal al banco central y la pérdida del señoraje”, *Revista Economíaunam*, vol. 10, no. 30, Economics Faculty-UNAM, Mexico, pp. 58-74.

Mélard G, & J. M., Pasteels., (2000). “Automatic ARIMA Modeling Including Intervention, Using Time Series Expert Software”. International Journal of Forecasting, 16, 497-508.

Ord K, & S. Lowe. (1996). “Automatic Forecasting”, The American Statistician, 50(1), 88-94.

Pfaff, B. (2015). “Econometric Time Series models: Part 1” Risk & Reward, Research and investment strategies, 20(2), 20-25

Reilly D. (2000). The Autobox System." International Journal of Forecasting, 16(4), 531-533.

Wuertz, D.,(2009). fExtremes: Rmetrics – Extreme Financial Market Data. R package version 2100.77. http://CRAN.R project.org/package=fExtremes.

Yule, G. U., (1927) "On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers", Philos. Trans. Roy. Soc. London Ser. A 226, pp. 267–298.

Codigo de Programación

```
#Titulo: Pronóstico de la Base Monetaria de México. Algortimo para la
  metodología Box-Jenkins
#Programa R por Eduardo Rosas Rojas; Última revisión: Abril 2017.
library(urca); library(forecast); library(zoo); library(lmtest);
  library(sandwich);library(fUnitRoots); library(tseries)
#------------------------------------------------------------------------------
# Importación de datos y transformación.
rm(list=ls())
print(paste("Ejecutado en: ", date()))
BM<-read.csv(file.choose(), header=T)
attach(BM)
bm<-ts(BM[,2],start=c(1985,12),freq=12)
#Se ha guardado la serie de tiempo, se reproduce con c: BM.csv
#------------------------------------------------------------------------------
#Paso 1: Análisis de Estacionariedad y Estacionalidad
# Descomposición de la serie en sus componentes :  tendencia,
  estacionalidad y ruido.
plot(stl((bm), s.window=4, t.window=12))
# Correlograma simple y correlograma parcial
tsdisplay(bm, lag.max=60, main = "Correlograma Base Monetaria")
#Prueba de Raíz unitaria
urdfTest(bm, lags = 1, type = c("nc"), doplot = T)
urdfTest(bm, lags = 1, type = c("c"), doplot = T)
urdfTest(bm, lags = 1, type = c("ct"), doplot = T)
#Identificación del número de diferencias ordinarias
ndiffs(bm, alpha=0.05, test=c("adf"), max.d=2)
# Identificación de diferencias estacionarias.
nsdiffs(bm, m=12, test=c("ocsb"), max.D=2)
#Transformación de diferencias logarítmicas y sus respectivas pruebas
dlbm<-diff(log(bm))
plot(stl((dlbm), s.window=4, t.window=12))
tsdisplay(dlbm, lag.max=60, main = "Correlograma Diferencia(log(Base
  Monetaria))")
urdfTest(dlbm, lags = 1, type = c("nc"), doplot = F)
urdfTest(dlbm, lags = 1, type = c("c"), doplot = F)
urdfTest(dlbm, lags = 1, type = c("ct"), doplot = F)
#------------------------------------------------------------------------------
#Paso 2:  Identificación del proceso ARIMA
# R cuenta con la función "auto.arima" del paquete "forecast" la cual
  selecciona automáticamente el orden del proceso ARIMA minimizando el
  Criterio AIC y/o BIC
arima_x=auto.arima(bm, max.p=12, max.q=12, max.P=12, max.Q=12,
  max.d=2, max.D=1, start.p=2, start.q=2, start.P=1,
  start.Q=1,stationary=FALSE , seasonal=FALSE,
  ic=c("bic"),stepwise=TRUE, test=c("kpss"), seasonal.test=c("ocsb"))
arima_x
# "auto.arima" indica un proceso SARIMA (0,1,2)X(0,1,3)[12]
#------------------------------------------------------------------------------
# Paso 3: Estimación del modelo (modelos)
# Usamos el comando "Arima "para realizar inferencia estadística; se
  especifica el modelo obtenido en el Paso 2, y otros dos modelos
  propuestos.
model1 <- summary(bm.arima<-Arima(bm, order=c(0,1,2),
  seasonal=c(0,1,3),include.mean=T, include.drift=F, include.constant=F,
  method=c("ML")))
model2<-summary(bm.arima1<-Arima(bm, order=c(0,1,2),
  seasonal=c(0,1,2),include.mean=T, include.drift=F, include.constant=F,
  method=c("ML")))
model3<-summary(bm.arima2<-Arima(bm, order=c(0,1,2),
  seasonal=c(0,1,4),include.mean=T, include.drift=F, include.constant=F,
  method=c("ML")))
#------------------------------------------------------------------------------
#Paso 4: Verificación del Modelo.
#Se grafican los residuos del modelo óptimo, sus correlograma Simple y
  Parcial, se aplica un a prueba de raíz unitaria, y otra prueba de No
  autocorrelación.
residuos<-(residuals(bm.arima))
tsdisplay(residuos)
urdfTest(residuos, lags = 1, type = c("nc"), doplot = F)
urdfTest(residuos, lags = 1, type = c("c"), doplot = F)
urdfTest(residuos, lags = 1, type = c("ct"), doplot = F)
Box.test(residuos, lag=1, type="Ljung")
Box.test(residuos, lag=2, type="Ljung")
Box.test(residuos, lag=3, type="Ljung")
Box.test(residuos, lag=4, type="Ljung")
Box.test(residuos, lag=1, type="Box")
Box.test(residuos, lag=2, type="Box")
Box.test(residuos, lag=3, type="Box")
Box.test(residuos, lag=4, type="Box")
# Se sigue el mismo procedimiento para los otros dos modelos
#------------------------------------------------------------------------------
# Paso 5: Pronóstico
# Pronosticamos 3, 6, y 9 meses (método recursivo) con intervalos al 95%
  de confianza. Se grafican los diferentes pronósticos
```


RINOE® All rights reserved.

```
bmarimaf3 = forecast(bm.arima, h=3, level=c(95))
bmarimaf6 = forecast(bm.arima, h=6, level=c(95))
bmarimaf9 = forecast(bm.arima, h=9, level=c(95))
par(mfrow = c(3,1))
plot(bmarimaf3, main = "Horizonte=3")
plot(bmarimaf6, main = "Horizonte=6")
plot(bmarimaf9, main = "Horizonte=9")
par(mfrow= c(1,1))
plot(bmarimaf9, ylab='Base Monetaria (Millones de Pesos)', xlab='Año' ,
 type='l', lwd="3", main = "Pronostico a 9 meses")
grid(lty=1, col=gray(.9))
summary(bmarimaf9)
# Se obtiene el pronóstico puntual y por intervalo.
#-----------------------------------------------------------------------------
#Paso 6. Validación del Pronóstico
#Se aplica la prueba Diebold-Mariano para identificar el modelo que
 proporciona el pronóstico óptimo.
f1<-bm.arima<-Arima(bm, order=c(0,1,2),
 seasonal=c(0,1,3),include.mean=T, include.drift=T, include.constant=F,
 method=c("ML"))  #Dado que el comando auto.arima nos señala que el
 modelo tiene deriva, colocamos "include.drift=T"
f1p<-forecast(bm.arima, h=9, level=c(95))
f2<-bm.arima1<-Arima(bm, order=c(0,1,2),
 seasonal=c(0,1,2),include.mean=F, include.drift=F, include.constant=T,
 method=c("ML"))
f2p<-forecast(bm.arima1, h=9, level=c(95))
f3<-bm.arima1<-Arima(bm, order=c(0,1,2),
 seasonal=c(0,1,4),include.mean=F, include.drift=F, include.constant=T,
 method=c("ML"))
f3p<-forecast(bm.arima2, h=9, level=c(95))
accuracy(f1p)
accuracy(f2p)
accuracy(f3p)
dm.test(residuals(f2p),residuals(f1p),h=9, alternative=c("greater"),
 power=2)
dm.test(residuals(f2p),residuals(f3p),h=9, alternative=c("greater"),
 power=2)
dm.test(residuals(f1p),residuals(f3p),h=9, alternative=c("greater"),
 power=2)
summary(bmarimaf9)
#---------------------------------------------------------------------------
```

ISSN-On line: 2524-2040
RINOE® All rights reserved.