COMPUTATIONAL METHODS IN THE IDENTIFICATION OF FORECASTING TIME SERIES MODELS

Ruiz-Gándara, A[1]

Caridad y Ocerin, J.M.[2]

Abstract

Most of the econometric software use goodness of fit measures as a tool to identify an ARIMA model, although the usual objective is forecasting. An alternative is to use predictive criteria for identification purposes. Both methods are compared with the aim of obtaining optimum forecasts. Some statistics such as BIC and others, obtained within the sample range are used together with some smoothing predictive criteria. They are compared as a tool for automatic selection of a useful model; the classical methods identify the correct model in just a third of the cases, as it is show with simulated series, and, in many cases a different model produces better forecasts. A case study with exports of the Spanish economy is presented with the comparative results attained with both approaches.

Keywords: automatic identification, Arima, predictive criteria, exports

1. INTRODUCTION

Analysing a time series requires formulating a reasonable guess about the underlying generating process of the data. This is a challenging process where uncertainty can not be fully removed, affecting the accuracy of forecasts. The usual approach is to employ different theoretical methods based in goodness of fit measures, some obtained from the likelihood function, such as BIC, AIC, and so on, or on residual based measures, as the determination coefficients or the MSE. Nowadays, in many practical situations, it is the forecasting accuracy witch is the central objective, and, it can even be tolerable to use an over o underparametrized model if it produces better forecasts. Dangl and Halling (2006) consider that only time-varying coefficients can produce consistent estimates of future values, while Roy, Somnath, Indrani and Kunal (2010) rely on predictive determination coefficients. ARIMA models are an easy to use tool for short term forecasting. This classic way of dealing with problems of time series modeling was developed by Box y Jenkins (1970). Many other authors have extended this

[1] Prof. Departamento de Estadística, Organización de Empresas, Universidad de Córdoba

[2] Catedrático de Estadística y Econometría, Universidad de Córdoba, ccjm@uco.es

procedure to solve the existing limitations in each particular case as stated by DeLurgio(1998), Zhang (2003) among others. In business applications, multiplicative ARIMA(p, d, q)×(P, D, Q)$_s$, are commonly used, although some non-linear models are common in financial time series. More developed models include neural networks or hybrid models as ARIMA and ANN, as in Wang, Zou, Su and Li (2013) and in Caridad, Sterba and Rubliková (2011), or mixtures of forecasts, has been used with good results.

Artificial neural networks are gaining increasing importance in many real applications, Allende (2002). The empirical results show that under certain conditions, even the simple, Zhang, Patuwo and Hu. (1998). The first neural network model was proposed by McCulloch and Pitts (1943), other models have been studied after, Elman (1990). Neural networks have found many successful applications, Santana (2006), Patiño (2012). In relation to the problem at hand, are frequently the use of neural networks forecasting nonlinear time series, Katijani, Hipel and McLeod (2005). Some authors have developed new models ARIMA, for example, Contreras (2009) and neural networks, Santana (2006) and even hybrid models, Caridad *et al* (2011), Zhang (2001) and Tseng *et al* (2002).

Here the focus will be on identifying ARIMA models, proposing a specification algorithm that could produce an optimum forecasting model, and such as that it could be imbedded in time series software with a computational efficiency. The problem can be stated in a simple way: given a particular time series S = $\{y_t;\ t = 1, 2,\dots, n\}$, and a forecasting range for $t = n + 1, n + 2, \dots, n + h$, and a set M of m ARIMA models, the purpose is to select a model that produces 'optimum' forecasts. This can be generalized at selecting a subset of M that could be used in generating a mixture of forecasts. A usual way of selecting a model is based on some goodness of fit measure obtained with S, although, as it can be show, it is not always the best choice for forecasting. To evaluate the criteria for identifying forecasting models, it is necessary to compare the real, $S_F = \{y_{n+1}, y_{n+2},\dots, y_{n+h}\}$. and the forecasted future values, $S_{F\hat{Y},i} = \{\hat{y}_{n+1,I}, \hat{y}_{n+2,i}, \dots, \hat{y}_{n+h,i}\}$, obtained with the i model in M. A general predictive measure can be expressed as a function of both sets

$$D_{Fi} = d_F(S_F, S_{F\hat{Y},i})$$

and the selection of an optimal forecasting model can be achieved minimizing it for $i = 1, 2, \dots, m$. This general measure could also be applied within the sample range. $D_i = d_F(S, S_{\hat{Y},i})$, but, this would not lead to optimum forecasts, as we shall prove.

In a real case, only the S data are available, and this set has to be divided in a training subset, S_Y. with $T = n - h$ data, using neural network terminology, and a forecasting set, S_{FY}, with the remaining h observations. The corresponding estimated and forecasted set of data will be labelled as $S_{\hat{Y}}$ and $S_{F\hat{Y}}$.

A predictive measure could be the mean square error, obtained with the forecasting errors, $a_{t,Mi} = y_t - \hat{y}_{t,i}$, associated to model i is

$$d_F(S_{FY}, S_{F\hat{Y}}) = MSE_{FM_i} = \frac{1}{h}\sum_{t=T+1}^{T+h} a^2_{t\,M_i}$$

and the corresponding goodness of fit is obtained with the first T data

$$d(S_Y, S_{\hat{Y}}) = MSE_{M_i} = \frac{1}{T - m_i} \sum_{t=m_i+1}^{T} a_{t\,M_i}^2$$

being m_i the data lost due to autoregressive and difference lags in M_i. Some alternative classical measures could be the *MAD* or the *RMAD*, and both could be calculated using the training or the forecasting sets. Another approach could be considered using moving training and forecasting sets of data, to check the stability of the model selection.

2. WEIGHTED GOODNESS OF FIT AND PREDICTIVE MEASURES

Additional measures both of goodness of fit and predictive statistics are obtained weighting the errors, with a monotone function, $w(n, t)$. The aim is to enhance the importance of the latest deviations. These alternative measures would be

$$d_w(S_Y, S_{\hat{Y}}) = MSE_{wM_i} = \frac{1}{T - m_i} \sum_{t=m_i+1}^{T} w(T, t) a_{t\,M_i}^2$$

$$d_{wF}(S_{FY}, S_{F\hat{Y}}) = MSE_{wFM_i} = \frac{1}{h} \sum_{t=T+1}^{T+h} w_F(T + h, t) a_{t\,M_i}^2$$

It is thus possible to assign larger weights to data nearer to the last data, that is to $t = T$, and decreasing values when going farther towards the past. For predictive measures, many situations of interest can arouse: larger weights for forecasts near to T, or to some particular forecasts, or even, it is possible to take into account the sign of the forecasting errors, or, alternatively, an aggregated forecast error over a period, such as a fiscal year. Many business situations can justify different weight functions: for example, in press distribution, forecasts have to be establish for $T + r + 1$, as there is a gap between the last data available at day $t = n$ is y_{t-r}, and the forecasts have to be made for the next day $n + 1$; or in sales forecasting, more importance is usually attributed to negative deviations than to overshooting the company's targets. In some situations, with monthly data forecasts, the objective is an aggregated measure over a fiscal year.

Some simple weight function could be based on a smoothing parameter, $0 < \alpha < 1$, linked to the importance of the errors regarding to their proximity to the last data available

$$d_w(S_Y, S_{\hat{Y}}) = MSE_{w(\alpha)M_i} = \frac{1}{T - m_i} \sum_{t=m_i+1}^{T} \alpha^{T-t} a_{t\,M_i}^2$$

$$d_{wF}(S_{FY}, S_{F\hat{Y}}) = MSE_{w(\alpha)FM_i} = \frac{1}{h} \sum_{t=T+1}^{T+h} \alpha^{t-T} a_{t\,M_i}^2$$

An 'optimal' model could be obtained with the minimization of the goodness of fit, or, alternatively, of the predictive measure. Nevertheless, models with different values of α could not be compared unless some normalization procedure between models is applied. For example, normalising the weights such as the could add to 1,

$$d_w^*(\mathrm{S}_Y,\mathrm{S}_{\hat{Y}}) = MSE^*_{w(\alpha)\mathrm{M}_i} = \frac{1}{\sum_{t=m_i+1}^{T}\alpha^{T-t}}\sum_{t=m_i+1}^{T}\alpha^{T-t}a^2_{t\,\mathrm{M}_i}$$

$$d_{w\mathrm{F}}^*(\mathrm{S}_{\mathrm{F}Y},\mathrm{S}_{\mathrm{F}\hat{Y}}) = MSE^*_{w(\alpha)\mathrm{FM}_i} = \frac{1}{\sum_{t=1}^{h}\alpha^{t-1}}\sum_{t=T+1}^{T+h}\alpha^{t-T-1}a^2_{t\,\mathrm{M}_i}$$

In both cases, the α value ought to be not far from 1, to avoid giving to much importance to the last error in the training set, or the first error in the forecasting set.

Similar measures could be based on mean absolute errors

$$d_w^*(\mathrm{S}_Y,\mathrm{S}_{\hat{Y}}) = MAE^*_{w(\alpha)\mathrm{M}_i} = \frac{1}{\sum_{t=m_i+1}^{T}\alpha^{T-t}}\sum_{t=m_i+1}^{T}\alpha^{T-t}\left|a_{t\,\mathrm{M}_i}\right|$$

$$d_{w\mathrm{F}}^*(\mathrm{S}_{\mathrm{F}Y},\mathrm{S}_{\mathrm{F}\hat{Y}}) = MAE^*_{w(\alpha)\mathrm{FM}_i} = \frac{1}{\sum_{t=1}^{h}\alpha^{t-1}}\sum_{t=T+1}^{T+h}\alpha^{t-T-1}\left|a_{t\,\mathrm{M}_i}\right|$$

To avoid problems with the selection of the smoothing parameter, an alternative procedure to select a model is to obtain, for each α value, the optimum specification, minimizing the goodness of fit or de predictive criteria

$$\underset{i=1..m}{\mathrm{mim.}} MSE_{w(\alpha)\mathrm{M}_i} = MSE_{w(\alpha)\mathrm{M}_{i*}} \qquad \underset{i=1..m}{\mathrm{mim.}} MSE_{w(\alpha)\mathrm{FM}_i} = MSE_{w(\alpha)\mathrm{FM}_{i*}}$$

and with the model $i*$ selected, obtain the goodness of fit or predictive measure (it could be the *MSE* or and alternative measure, $\mathrm{MSE}_{\mathrm{M}i*(\alpha)}$, for different values $\alpha \in [a, 1]$, and, finally, obtaining the selected model that

$$\underset{\alpha\in[a,1]}{\mathrm{mim.}} MSE_{\mathrm{M}_{i*(\alpha)}} = MSE_{w(\alpha)\mathrm{M}_{i*(\alpha*)}}$$

Different weight structures can be proposed taking into account the objectives of the forecasting process. For example a water company may need monthly forecast of sales, but with an objective set for the fiscal year, and in these cases some additional weights must be considered. While in some other cases, the forecast precision is measured taking into account only the aggregated deviation in a full fiscal year.

The usual goodness of fit measures are based either on the residual sum of squares, or on the likelihood function (witch depends on normal assumptions about the distribution of the error structure, although, in this case, both are related). In a model with k parameters (in ARIMA models without constant term, $k \leq p + q + Ps + Qs$), these measures are of the type

$$IC = \ln \hat{\sigma}_a^2 + k\frac{c(T)}{T}$$

been $\hat{\sigma}_a^2$ the variance of the model white noise, k, the number of parameters in the model, and $c(T)$ a 'penalty' imposed upon the final result based on the number of estimated parameters. Some usual goodness of fit measures are the Akaike information criteria with $c(T) = 2$, the Bayesian information criteria, Schwartz (1978) with $c(T) = \ln T$, and some others. In some computer packages the obtained values can differ, and it is usual to find the goodness of fit based on the MSE, as in Forecast Pro

$$AIC = \ln MSE + \frac{k}{T} \qquad BIC = \ln MSE + \frac{k \ln T}{2T}$$

or in the R-project

$$AIC = T\ln \hat{\sigma}_a^2 + T + T\ln 2\pi + 2k \qquad BIC = T\ln \hat{\sigma}_a^2 + T + T\ln 2\pi + 2k\ln T$$

The Schwartz criteria is preferred to the *AIC* to compare different specifications, although with small samples, this is not always the case, Lütkepohl (1991). And, one has to take into account that the correct model (it can be known in Monte Carlo experiments) is not always the best for forecasting purposes, Sánchez and Peña (2001).

Many authors have treated alternatives to these information criteria, as in Sugiura(1978), Ogata(1980), Chow (1981), Stone (1982), Findley (1985), Bozdogan (1987), Shibata (1989), Cavanaugh (1997) or Burnham and Anderson (2002), for the *AIC*, or Chow (1981), Haughton (1989), Raftery, Madigan y Volinsky (1995) and Cavanaugh and Neath (1999), in the case of the *BIC*.

Others measures of the predictive quality are Theil (1958 and 1966) indices.

$$U_1 = \frac{\sqrt{MSE}}{\sqrt{\frac{1}{h}\sum_{t=1}^{h} y_{n-h+t}^2} + \sqrt{\frac{1}{h}\sum_{t=1}^{h} \hat{y}_{n-h+t}^2}} \qquad U_2 = \frac{\sqrt{MSE}}{\sqrt{\frac{1}{h}\sum_{t=1}^{h} y_{n-h+t}^2}}$$

The mean square error (*MSE*) can be disaggregated in the well known bias, variance and covariance components

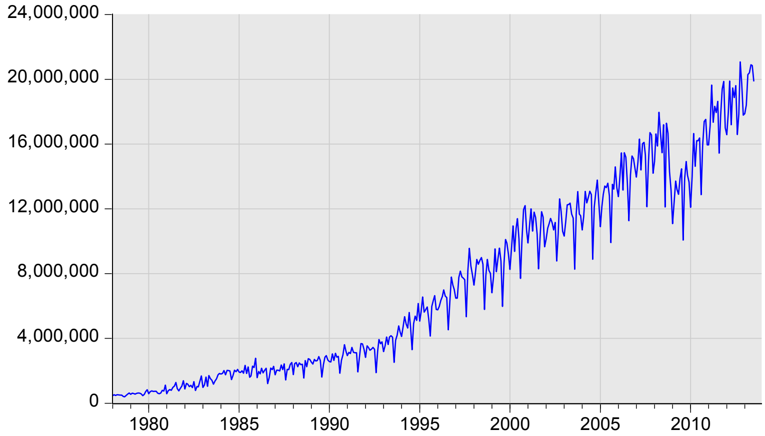
$$MSE = \frac{1}{h}\sum_{t=1}^{h}(y_{n-h+t} - \hat{y}_{n-h+t})^2 = \left(\bar{\hat{y}} - \bar{y}\right)^2 + \left(s_{\hat{y}} - s_y\right)^2 + 2(1 - r_{\hat{y}y})s_{\hat{y}}s_y$$

and the decomposition of the Theil component can be used as a measure of what can be expected in the forecasting power. Specially, in these times of crisis, it is not unusual to find a large bias component in the forecasting set $S_{F\hat{Y},i}$

3. AUTOMATIC IDENTIFICATION OF MODELS IN A CASE STUDY

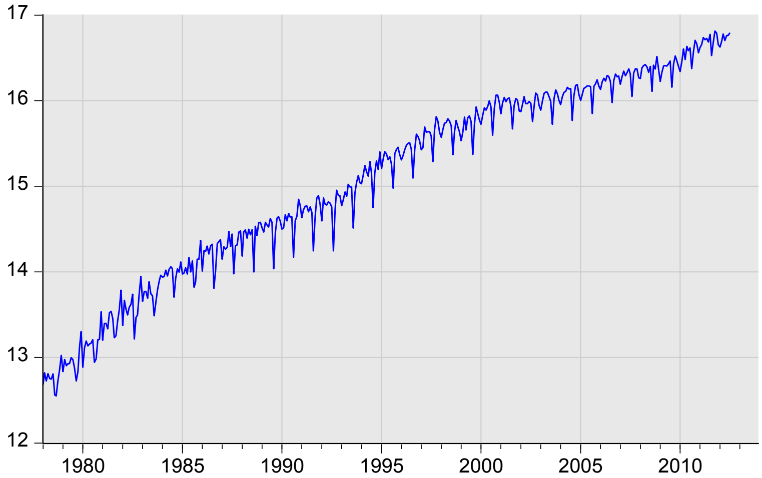
The export sector in Spain has not been able to cover the country's imports until very recently. Its evolution shows a continuous growth accelerated in the mid nineties. In 2008 the global financial crisis was felt with a significant drop in sales, only to be recovered since 2010, as can be seen in the following figure.

Figure 1. Monthly exports in 10^3 euros



Import figures have been growing, although at a lower rate, due to the fall in internal consumption associated to internal fiscal crisis. Spain needs to balance, in the next few years, its public sector finances, as, both in the Administrations (local, regional and national), and in the Social Security, the deficits are causing a lack of credit to the private sector and a stagnated consumption. The fall in real wages and the increase in productivity are already showing positive signs, reflected by the sharp increase in the stock market and in the creation of new companies. This change of trend is conditioned by political factors and regional tensions, as well as the chronic public deficit that is far from been solved. The importance of the evolution of exports are consequences of both the situation in the internal market and the increase of competitiveness of the Spanish firms. The short and medium term risk premium has been falling recently, and the evolution of the external sector is part of the whole picture. It is convenient, thus to be able to forecast for the next fiscal year the evolution of the export markets; of course, we are not taking with this approach, some exogenous factors that could influence this trend, and the basic hypothesis is that the external sector will be stable in the short term.

Figure 2. Linearized log-export series



This case study carries out an application to analyze the evolution of monthly Spanish exports in the last 35 years. The data available are the monthly exports (Figure 1 from 1978 to July 2013. To specify a model, the first $T = 415$ data are considered as the training set, and the last $h = 12$ defined as the objective of forecasting. The set of possible models considered, M, include ARIMA specifications with some non practically restrictive conditions with $p + q \leq 4$, $P + Q \leq 4$, $d \leq 1$, and $D \leq 1$, that is 324 possible models, with autorregresive and moving average polynomials with o without all their central terms. Previously, some linearization has been carried out, with an intervention to take into account the temporary change in the level of the series due to the 2008-9 crisis, and also to remove the calendar effects. A logarithmic transformation has also been applied. The resulting series is show on the figure 2.

To identify an optimal model, the different criteria developed in the previous section, as well as the classical criteria, are employed, comparing the obtained results. An automatic procedure to identify an ARIMA model is include in some econometric packages, as Forecast Pro, Autobox, or in the Eurostat sponsored Demetra+; but all of them rely on, either minimizing the BIC criteria, or on a set of statistical validation tests, applied as a rule based decision system. In practical work with business data, our experience shows that these procedures need to be strictly supervised by an expert, or they could lead, sometimes, to awkward results.

To select a model using the former criteria considered, an optimum for each of them is obtained, leading to the following ARIMA specifications

Table 1. Classical criteria in the identificación on an ARIMA model

AIC	***BIC***	U_1	U_2
$(1,1,2)\times(2,0,2)_{12}$	$(0,1,2)\times(1,0,1)_{12}$	$(1,0,2)\times(2,1,2)_{12}$	$(1,0,2)\times(2,1,2)_{12}$

while, when using the weighted goodness of fit and forecasting criteria, the corresponding specifications are, using different values for the smoothing constant

Table 2. Smoothly weighted criteria in the identification process

A	$MSE_{w(\alpha)}$	$MSE_{w(\alpha)F}$	$MAE_{w(\alpha)}$	$MAE_{w(\alpha)F}$
0.95	$(2,1,0)\times(1,0,2)_{12}$	$(1,0,0)\times(2,0,1)_{12}$	$(2,1,0)\times(1,0,2)_{12}$	$(1,0,0)\times(2,0,2)_{12}$
0.96	$(2,1,0)\times(1,0,2)_{12}$	(1,0,0)×(2,0,1)12	$(2,1,0)\times(1,0,2)_{12}$	$(1,0,0)\times(2,0,2)_{12}$
0.97	$(2,1,0)\times(1,0,1)_{12}$	$(1,0,0)\times(2,0,1)_{12}$	$(2,1,0)\times(1,0,1)_{12}$	$(1,0,0)\times(2,0,2)_{12}$
0.98	$(2,1,0)\times(1,0,1)_{12}$	$(1,0,0)\times(2,0,2)_{12}$	$(2,1,0)\times(1,0,1)_{12}$	$(1,0,0)\times(2,0,2)_{12}$
0.99	$(1,1,2)\times(1,0,1)_{12}$	$(1,0,0)\times(2,0,2)_{12}$	$(1,1,2)\times(2,0,2)_{12}$	$(1,0,0)\times(2,0,2)_{12}$

It can be seen that the weighted criteria tend to produce different specifications, but the identified models are only slightly dependent on the smoothing constant.

To select the optimal model, the minimum value in each of the criteria is considered. Their respective residuals are represented in Table 3.

Table 3. Results with different identification criteria

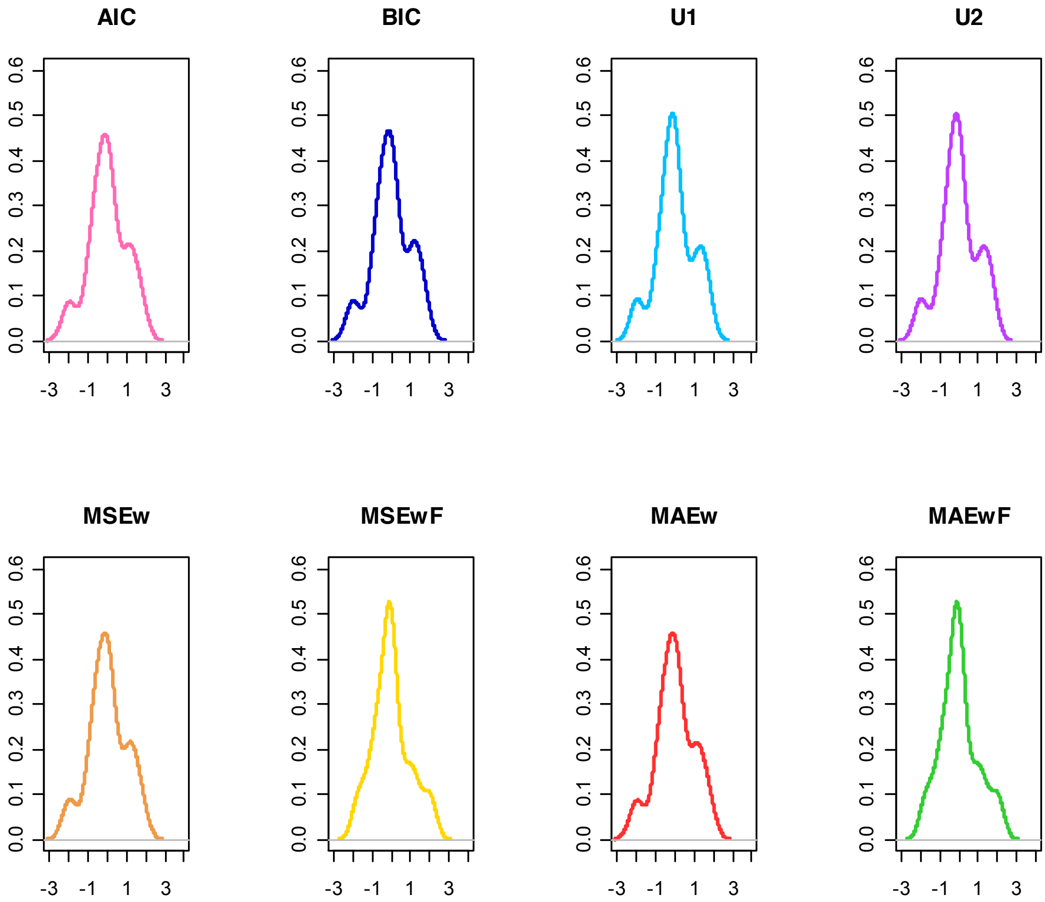
AIC	*BIC*	*U* **Theil**	*MSE*	MSE_F
0.04137887	0.04252467	0.04307843	0.04353314	0.04307843
$(1,1,2)\times(2,0,2)_{12}$	$(0,1,2)\times(1,0,1)_{12}$	$(1,0,2)\times(2,1,2)_{12}$	$(1,0,2)\times(2,1,1)_{12}$	$(1,0,2)\times(2,1,2)_{12}$

Using different values of the smoothing constant, the identification process can produce different results, although they are quite stable, as can be seen in table 4.

Table 4. *MSE* and *MAE* measures with different specifications

α	$MSE_{w(\alpha)}$	$MSE_{w(\alpha)F}$	$MAE_{w(\alpha)}$	$MAE_{w(\alpha)F}$
0.95	0.04455343	0.04047769	0.04455343	**0.04033697**
0.96	0.04455343	0.04047769	0.04455343	**0.04033697**
0.97	0.04357228	0.04047769	0.04357228	**0.04033697**
0.98	0.04357228	**0.04033697**	0.04357228	**0.04033697**
0.99	**0.04129497**	**0.04033697**	**0.04137887**	**0.04033697**

Figure 3. Distributions of the residuals for each specification



To appreciate the differences between the actual and the forecasted values with each of the identified models, the corresponding plots can be seen in figures 4 and 5

Figure 4. Forecasted and actual values with the log-linearized series

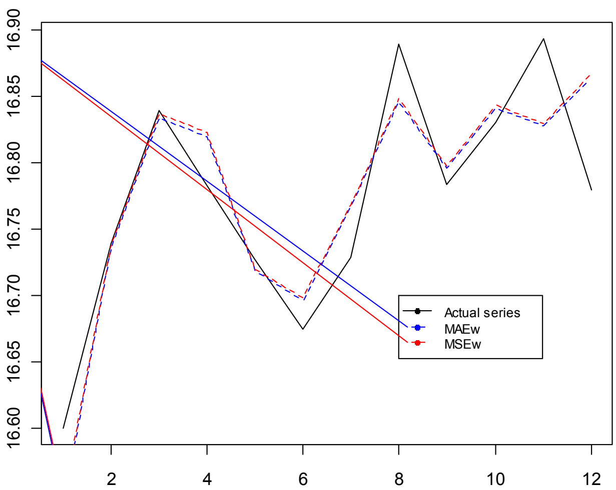
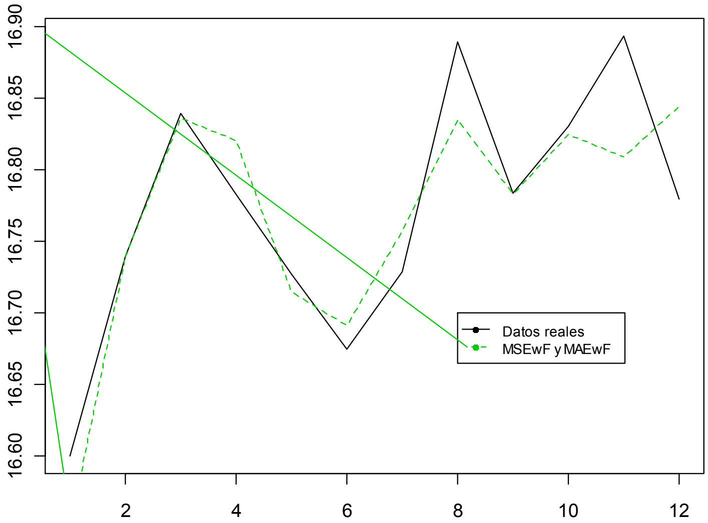


Figure 5. Forecasted and actual values with the log-linearized series



Using the original data, the goodness of fit and the predictive measures behave in a similar way, as can be seen in the following tables

Table 5. Results on the original series with different identification criteria

AIC	***BIC***	***U* Theil**	***MSE***	***MSE*F**
814700.7	838190.6	834387.1	844454.8	834387.1

Table 6. *MSE* and *MAE* measures on the original series with different specifications

α	$MSE_{w(\alpha)}$	$MSE_{w(\alpha)F}$	$MAE_{w(\alpha)}$	$MAE_{w(\alpha)F}$
0.95	878911.3	805462.2	878911.3	**802319.7**
0.96	878911.3	805462.2	878911.3	**802319.7**
0.97	859244.9	805462.2	859244.9	**802319.7**
0.98	859244.9	**802319.7**	859244.9	**802319.7**
0.99	**809859.6**	**802319.7**	**814700.7**	**802319.7**

An additional procedure to validate the forecasts is the well-known Theil decomposition of the mean square error in bias, variance and covariance components. With the model selected using different criteria, it can be seen that the proposed weighted and weighted predictive criteria tend to produce lower values in the bias component, although, in some case the variance component is not the lowest possible. Table 7 summarizes the results.

Table 7. Theil's *U* decomposition for different models

Identification criteria	Bias component	Variance component	Covariance component
AIC, MAE_w	4.304423E-6	5.184611E-7	0.0003995075
BIC	2.498928E-6	1.497917E-6	0.0004008668
U	2.600339E-6	8.616724E-6	0.0004129558
$MSEw$	7.441332E-7	5.058581E-7	0.0003989647
MSE_{wF}, MAE_{wF}	7.129798E-7	4.026716E-6	0.0003851257

4. CONCLUSIONS

The proposed identification of ARIMA models is based on goodness of fit and on prediction criteria, introducing a weight function to take into account the different importance attributed in practice to the forecasting errors. To identify a model with a particular time series, the proposed criteria enhance the predictive ability on the selected model over the classical measures in the literature. A smoothing weight scheme is computationally easy to implement and reproduce the intuitive perception about the importance of the errors, importance that decreases when the temporal instant is farther away from the last or actual value. The smoothing constant, α, should be close to one, as the selection of models would be based on the more recent errors, with the corresponding loss of information; thus a range between 0.95 and 0.99 is suited in most of practical situations. Of course, different weight structures can be used to accommodate the aims of the forecasting process, such as obtaining more precise aggregate forecast over a fiscal year, or enhance the importance of a temporal range of forecast errors, as can be the case in press distributions. Most of the commercial software available use the *BIC* or *AIC* criteria or a set of statistical based rules to identify the 'best' model. The proposed MSE_w or MSE_{wF} measures tends to produce lower forecasting errors. using *MSE* or *MAE* aggregate measures. using the information criteria to select the forecasting model. In the case study presented, the monthly exports of Spain time series is analyzed, and with a forecast horizon of twelve months, the proposed criteria confirm these results. Previously some business time series, as water demand in urban areas, have led to similar results. Comparing the *BIC* criterion with the proposed *MSE_w* method, there is again an improvement when using the later. In term of goodness of fit in the training set, the *BIC* criterion achieved better results as could be expected. With simulated data the results are similar, but the *BIC* criterion is more suited to detect the true process generating the time series, while for forecasting purposes, the proposed identification procedures tends to produce more accurate results.

5. REFERENCES

Akaike, H.: (1973), Information theory and an extension of the maximum likelihood principle, *Akademiai Kiado*, 1, 267-281.

Allende, H., Moraga, C. and Salas, R.: (2002), Artificial Neural Networks in the Time Series Forecasting; A Comparative Analysis, *Kybemetik*, 38, 685-707.

Box, G.E.P. and Jenkins, G.M.: (1970), *Time Series Analyses Forecasting and Control*, San Francisco: Holden Day.

Bozdogan, H.: (1987), Model selection and akaike's information criterion (aic): "The general theory and its analytical extensions, *Psychometrika*, 52, 345-370.

Burnham, K. P., and Anderson, D. R. :(2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer Verlag.

Caridad y Ocerín, J.M, Ján Sterba and Rublíková, E.: (2011), *Hybrib ARIMA-Neural Network Model for Prediction of Water Consumption Aggrete*, International Conference on Social Sciences and Society.

Cavanaugh, J. E. and Shumway, R. H.: (1997), *A bootstrap variant of AIC for state space model selection*, *Statistica Sinica*, 7(2), 473-496.

Cavanaugh, J. E., and Neath, A. A.: (1999),Generalizing the derivation of the Schwarz information criterion, *Communications in Statistics-Theory and Methods*, *28*(1), 49-66.

Chow, G. C.: (1981), A comparison of the information and posterior probability criteria for model selection, 16(1), 21-33.

Contreras, J. and Idrovo, B.: (2009), "*Un modelo SARIMA para predecir la Tasa de Desempleo en Chile*".

Dangl, T. and Halling (2006), "*Predictive Regressions with Time-Varying Coefficents*", Social Science Research Network. WP

DeLurgio, S.A. :(1998), "*Forecasting Principles and Applications*", Tom Casson, New York.

Elman, J.L.: (1990), Finding structure in time, *Cognitive Science*, 14, 179–211.

Findley, D. F.: (1985), On the unbiasedness property of aic for exact or approximating linear stochastic time series models, *Journal of Time Series Analysis*, 6(4), 229-252.

Haughton, D.: (1989). Size of the error in the choice of a model to fit data from an exponential family, *Sankhyā: The Indian Journal of Statistics,* Series A, 45-58.

Katijani, Y., Hipel, W.K. and Mcleod, A.I.:(2005), Forecasting nonlinear time series with feedforward neural networks: A case study of Canadian lynx data, *Journal of Forecasting*, 24, 105–117.

Lütkepohl, H.: (1991), "*Introduction to multiple time series analysis",* Springer, New York,

McCulloch, W.S. and Pitts, W.: (1943), A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biology*, 5(4),115-133.

Ogata, Y. :(1980), Maximum likelihood estimates of incorrect markov models for time series and the derivation of AIC, *Journal of Applied Probability*, 17(1), 59-72.

Patiño, V.M.O.: (2012), "*Redes Neuronales Artificiales en las Ciencias Económicas"*, UN-RCE-CID.

Raftery, A.E., Madigan, D. and Volinsky, C.T. :(1995), Accounting for model uncertainty in survival analysis improves predictive performance, *Bayesian Statistics*, 5, 323-349.

Roy, P.P., Domnath P, Indrani, M. and Kunal R.: (2010), On two Novel Parameters for Validation of QSAR Models, *Molecules*, 15, 604-5.

Santana, J.C. :(2006), Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana, *Revista Colombiana de Estadística*, 29(1), 77-92.

Schwarz, G. :(1978), Estimating the dimension of a model, *Annals of Statistics*, 6(2), 461-464.

Shibata, R. :(1989), "*From data to model*, *in Statistical aspects of model selection*", pp. 215-240. Springer-Verlag New York, Inc., New York, NY, USA.

Stone, C. J. :(1982). Local asymptotic admissibility of a generalization of Akaike's model selection rule, *Annals of the Institute of Statistical Mathematics*, 34(1), 123-133.

Sugiura, N.: (1978), Further analysis of the data by Akaike's information criterion and the finite corrections, *Communications in Statistics - Theory and Methods*, 7(1), 13-26.

Tseng, F.M., Yu, H.C. and Tzeng, G.H.: (2002), Combining neural network model with seasonal time series ARIMA model, *Technological Forecasting & Social Change*, 69, 71–87.

Theil, H.: (1966), "*Applied economic forecasting"*, . 4. North-Holland Pub. Co.

Wang, Li, Haofei Zou, Jia Su and Ling, Li: (2013), An ARIMA-ANN Hybrid Model for Time Series Forecasting. *System Research and Behavioral Science*. 30(3), 244-259.

Zhang, G.P.: (2003), Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, 50, 159-175.

Zhang, G.P., Patuwo, E. B., and Hu, M. Y.: (2001), A simulation study of artificial neural network for nonlinear time series forecasting, *Computers & Operations Research*, 28, 381–396.

Zhang, G.P., Patuwo, E. B., and Hu, M. Y.: (1998), Forecasting with artificial neural networks: The state of the art, *Journal of Forecasting*, 14, 5.