COPYRIGHT

Luxembourg: Office for Official Publications of the European Communities, 2003

ISBN 92-894-5347-8
ISSN 1725-4825
Cat. No. KS-AN-03-019-EN-N

2003 EDITION

Large Scale Fitting of ARIMA Models and Stylised Facts of Economic Time Series

EUROPEAN COMMISSION

Europe Direct is a service to help you find answers to your questions about the European Union

New freephone number:

00 800 6 7 8 9 10 11

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server (http://europa.eu.int).

Luxembourg: Office for Official Publications of the European Communities, 2003

ISBN 92-894-5347-8
ISSN 1725-4825

TABLE OF CONTENTS

Large Scale Fitting of ARIMA Models and Stylised Facts of Economic Time Series

Björn Fischer* **Christophe Planas§**

February 1998

*Planistat Luxembourg, Bld. J.F. Kennedy, L-1855, Luxembourg.

§Joint Research Centre, Via E.Fermi, 1, TP 316, I-21020, Ispra, Italy. e-mail: christophe.planas@jrc.it . Please send correspondance to the second author. This paper was written within the framework of a study on Seasonal Adjustment methods, on request of EUROSTAT. The first author is acting as consultant for EUROSTAT, unit D1-Industrial Short Term Indicators, while the second author is related to unit A4-Research and Developement (contract nr. 6663002). The ideas expressed here are the authors' and do not necessarily reflect the position of EUROSTAT. The authors thank Berthold Feldmann, EUROSTAT-D1, for his interest in this work, and Raoul Depoutot, EUROSTAT-A4, for useful comments.

Abstract

The Statistical Office of the European Communities (Eurostat) publishes information on the economies of the Member States using, for some units, some model-based procedures to treat several features of economic time series. The quality of the information published is thus related to the capacity of these models, namely univariate Arima models with exogenous regressors, to adequately describe a vast majority of economic time series. We evaluate that capacity on a set of 13238 monthly series. The results of our experiment give several messages: 1) the sensitivity of different economic indicators to calendar events can be quantified; 2) the occurences and the typology of outliers found in practice are detailed; 3) information is obtained about the stationary behavior of the series; 4) the practical relevance of several model specifications can be evaluated; 5) the type of the mis-specifications found is detailed, yielding for example an indication on nonlinear patterns actually encountered in monthly series.

1 Introduction

One of the tasks of the Statistical Office of the European Communities (EUROSTAT) consists in making available information on the Member States economies. That information is subject to a statistical treatment which regards some particular features of economic time series. Namely, the trading days rythmn, the easter recess effect, some data irregularities, the series growth, and some unobserved movements like trend and seasonality are of main interest. In some units, all the related analysis is performed in a model-based framework through the use of the programs TRAMO-SEATS (see Gomez and Maravall, 1996). The methodology implemented is that of univariate regression with time series errors of the ARIMA-type (see for example Bell, 1995, Fuller, 1991, and Tsay, 1984), plus some developments related to outlier detection and correction and to a fully automised model identification procedure (see Gomez 1997). These last two advances were crucial for a massive model-based treatment of time series.

The quality of the information published is thus related to the capacity of univariate ARIMA models with exogenous regressors to describe a vast majority of economic time series. In this article, we evaluate that capacity on a set of 13238 monthly series. As far as we know, no results on such large scale fitting of stochastic linear models with exogenous regressors are available in the literature. Besides the overall capacity of these models in describing economic series, the results of our experiment give several messages. First, the sensitivity of different economic indicators to calendar events can be quantified. Second, the occurences and the typology of outliers found in practice are detailed. Third, information is obtained on the stationary behavior of the series. Fourth, the practical relevance of the airline model (see Box and Jenkins, 1976), which is used in many applied works, can be evaluated. Fifth, the type of the mis-specifications found is detailed, yielding for example an indication on nonlinear patterns actually encountered in monthly series.

We present in section 2 the data collection process, in section 3 the methodological treatment, and in section 4 we discuss the results of our experiment.

2 Data collection

The time series are taken from the Industrial Short-Term Indicator section of the Eurostat database "New Cronos". They comprise series of the 15 Member States of the European Union plus the European total and a few series from the United States and from Japan. For the classification of the industrial activities, the revised Classification of Economic Activities within the European Communities (Nace Rev.1) was used.

Five different areas are covered: industrial production, turnover, new-orders, import and export. The number of series in every group is 2512, 2206, 1641, 3547, 3332, respectively, for a total number of 13238 series. All the series are monthly. The samples sizes are roughly distributed as follows: 10100 series are of length in $[85, 105]$, 1500 in $[130, 155]$, 1300 in $[195, 212]$, the other sample lengths being roughly uniformly distributed outside these intervals between the minimum of 75 and the maximum of 212. A short description is developed in the rest of that section; a more detailed information can be found in Eurostat, 1997.

The production index measures the evolution in volume at constant prices of gross value added produced by an observation unit of a given activity. As most of the Member States only supply trading day adjusted production indexes, preliminary transformed data had to be used. The turnover index, which measures the turnover of the total of products and services invoiced by the observation unit, is measured in current value and was used in three presentations: domestic turnover, external turnover, and total turnover. New orders correspond to all orders received in the course of a reference month minus the cancellations that occur in this period. They were split up in the same way than the turnover index, and are also evaluated in current prices (value index). The indexes of imports and exports are divided into value and volume data. The trade data on industrial products of the Member States of the Eu (intra and extra Eu together) and of the Eu (only extra Eu trade) was used.

3 Methodology

Every series is treated separately with the program Tramo. The different steps of the treatment can be found in Gomez and Maravall, 1996; details are also given in Gomez, 1997. We briefly summarize the different tasks performed below.

A test for the log-level specification based on a range-mean regression is first computed. According to the result, the data are log-transformed or not. The airline model is then used to test for trading days and easter effect, and to compute a generalised least-squares prior correction if these are found significant. In our experiment, the trading days regressors have been specified as made up of 6 variables plus a length-of-month index, while an easter effect variable has been specified so as to describe an effect lasting 8 days (see, for example, Bell, 1995). Then a search for the differencing orders of the Arima model starts. The procedure is based on the results of Tiao and Tsay (1983) and of Tsay (1984). Broadly, autoregressive polynomials and Arma models are sequentially fitted to determine the number of autoregressive unit roots. Once the differencing orders have been selected, identification of the Arma model orders is performed on the basis the Bic criterion (see Hannan and Rissanen, 1982). The search puts the emphasis on low order and on balanced models; the model with maximum order that may be considered is $(3, 2, 3)(1, 1, 1)_{12}$. Estimation is computed by exact maximum likelihood using the algorithm in Mélard (1984).

An outlier detection and correction procedure is conducted along the lines of Chang et al. (1988) and of Tsay (1986), with some improvements. First, outliers are detected and corrected singularly and then a multiple regression is performed to eliminate spurious ones. In our experiment, three types of outliers have been considered: additive outlier (Ao), temporary change (Tc), and level shift (Ls). The critical value for outlier detection has been set at 3.5, 3.7 and 4.0 for series lengths less than 130 observations, between 131 and 180, and more than 180 observations, respectively (see Chang and al., 1988).

In order to check whether the resulting regression model with Arima errors have been able to adequately describe the properties of the series, an analysis of the residuals has been performed. The statistics considered to check whether the residuals are uncorrelated white noises are the Ljung-Box Q-statistics computed on the first 24 lags, and the Box-Pierce Q-statistics (denoted Q_s) computed on seasonal lags 12 and 24 (see Ljung and Box, 1978, and Pierce, 1978, respectively). Independency is verified by computing these statistics on the squared residuals (Mc Leod and Li, 1983). A further check is performed by comparing the skewness and kurtosis of the residuals distribution with the theoretical third and fourth moments of the normal distribution. Finally, all the models are re-estimated with 12 observations left apart for computing post-sample predictive tests (see Harvey, 1989, p.271).

The tests rejections are reported for nominal significance levels of 5% and of 1%. This last test-size is usually not considered in applied works focusing on few time series, but it makes sense considering low test-sizes when one is confronted to such a massive set of series. Under the null hypothesis of correct model specification, it would yield 130 series to analyse further, a number which is already costly. Considering instead the standard 5% level on every diagnostic for deciding which series needs a more accurate analysis could be difficultly affordable for a statistical department dealing with massive sets of series. We now turn to discussing the results.

4 Results

Table 1 shows the proportion of time series presenting a significant calendar effect by group of indicators. It can be seen that over the 13238 series of our sample, 35% are sensitive to the trading days rythmn and 14% to easter recess. The results concerning production indexes are however somewhat misleading. Normally suppliers provide EUROSTAT with trading day adjusted production indexes as most of the economic sectors are sensitive to the trading day rythmn. But some series may not have been subject to that treatment, or not in a satisfying way, and consequently a significative effect was found in 19% of the production series. Among the four other groups, turnover series are the most concerned by the trading days with an incidence found in 50% of the series. Turnover series are closely related to sales, so that this result is consistent with the general agreement that sales are highly affected by calendar events. One-third of new orders and international trade series are found sensitive to a trading days rythmn.

Table 2 shows the proportion of outliers found in every series, by type and by economic sector. The mean number of outliers found by series is 1.25, that is pretty low. Production is the sector where more outliers are detected (1.80 by series), while import series are those presenting the less data irregularities (.97 by series). For every group, the outlier-type the most often found is the AO: 40 % of series have at least 1 AO, against 25% for LS and slightly less for TC. The relatively large number of outliers in the production sector is due to the large proportion of AO: more than .5 by series. The proportion of LS and of TC is roughly similar among the different groups.

We now examine the ARIMA specifications that have been identified. Table 3 gives mean results, still displayed by groups of series. First it is seen that from 78% to 88% of the series need a prior log-transform. For the groups where the indexes are in current prices, this is consistent with the inflation effect that current value variables usually embody. But it was also needed for a vast majority of production indexes, which are measured in volume. Besides the log-transform, very few series show a stationary behavior: for only 4% of the series linear differencing was not necessary. In more that 70% of the cases, that non-stationary behavior needs both a regular and a seasonal difference ($\Delta\Delta_{12}$) to be corrected. Regular difference or seasonal difference on their own is sufficient only in very few cases (8% and 12%, respectively). Notice that the need for a two regular differences is nearly never met. Seasonal unit roots are present in more than 80% of the models fitted. Not much discrepancies between groups can be seen with respect to the stationarity properties.

Nearly 95% of the series of every sector are described by a seasonal model. Purely seasonaly or purely regular models were used in less than 5% of the cases, showing the importance of multiplicative models in describing the behavior of monthly economic indicators. Finally, 60% of the models considered were the airline model, 70% a pure IMA and 5% a pure ARI.

The diagnostic checks on the models fitted are detailed on table 4a. We first discuss the results obtained with a 5% nominal size of the tests, considering every test separately. It is seen that the Ljung-Box statistic points to uncorrelated residuals for more than 90% of the series. The Box-Pierce statistic designed to indicate some remaining correlations at seasonal lags show less than 4% departures from the white noise hypothesis, that is less than the nominal size. Regarding the residuals distribution, only 10% of the overall set of residuals show a significant departure from a normal one. Yet, that proportion reaches 24% for residuals kurtosis in industrial production indexes. It is the sector where more AO were found. This suggests that industrial production is subject to some irregularities. Another feature of interest concerns the new orders group, where 11% of the residuals show some asymmetry inconsistent with the normal hypothesis. In close agreement with the results about the residuals distribution, tests for residual independency show that roughly 10% of the series embody some significant nonlinear structure. The group the most concerned seems to be industrial production. All the departures found are not strongly evident: considering 1% nominal sizes for the tests lowers the proportion of rejection to 1% for correlation and normality statistics, to 7% for linear independency tests.

Combining the information yielded by the different diagnostics is useful to understand the overall acceptation level and the mis-specifications found. Table 4b displays the results, and it can be seen that for 65% of the series not any diagnostic is significant at the 5% level. International trade series are the best described by linear regression with ARIMA errors, satisfying any check at 5% in 72% of the series, while on the other hand production series are correctly described in 50% of the cases. Lowering at 1% the significance level of every test increases the number of acceptable models from 92% to 72% for these two groups which remain the extremes, and to 86% for the overall series. This result is pretty conforting about the capacity of ARIMA models with exogeneous regressors to describe economic time series. Furthermore, if the interest centers on models able to describe the series second moments, then it is seen on table 4b that 90% of the models fitted let residuals with white noise properties, 98% at the 1% level. As discussed in Planas (1998), that feature is most important when the aim of the analysis is to perform seasonal adjustment or trend extraction through optimal signal extraction. The overall result together the high proportion of models catching the correlation pattern of the series validates the methodological choice of an ARIMA-model-based approach for publishing trends and seasonally adjusted series that some units of EUROSTAT operated (see EUROSTAT, 1998).

The departures from correct model specification are also of interest. The distribution of 23% of the residuals present a distribution not in agreement with a normal distribution, of 7% at the 1% level. Nonlinear dependencies are evident in 18% of the residuals, nearly never related to the seasonal lags. That overall proportion is rather low, and mostly due to the production indexes where roughly 30% of the series present some evidences of nonlinearities. The relatively low proportion of models correctly describing production indexes was thus mainly due to nonlinearities in the data. Lowering the critical value for outlier detection would have mechanically rised the number of satisfying models found for these indicators.

It is interesting to evaluate the performances of some particular model specifications. Table 5 gives the performances of the airline model, of IMA models and of ARI models, in fitting the data. The airline model was introduced by Box and Jenkins, 1976, and it has been very widely used both in applied works and in methodological papers. It can be seen on table 5 that it has the appeal of correctly describing actual patterns since, at the 5% and 1% levels, for more than 40% and 50% of the 13238 series the diagnostic checks were satisfying. For IMA models which are more general, that proportion rises to 48% and 62%, respectively. ARI models were less often used, and so only few are found satisfactory in describing actual patterns. That is more related to the automatic model procedure of the software TRAMO which favors balanced and low-orders models rather than to the actual performances of autoregressive models. Notice that if the attention concentrates on second-order moments, then the airline model is able to deliver white-noise residuals in nearly 60% of the cases, which is definitively an impressive proportion.

Finally, table 6 show the results of post-sample predictive tests. The performances of different methods for forecasting economic time series has been the subject of a large debate in the time series literature; for an overview, see Fildes and Makridakis, 1995. We do not pursue in this direction here. Not only this would be outside our scope, but the methodology we have used which also performs automatic corrections for both outliers and calendar effects is more general that those typically involved in these forecasting competitions. Rather, we concentrate on the proportion of series which could be forecast in a consistent way using regression models with ARIMA errors. The results displayed in table 6 are actually very satisfying, given the relatively long in-sample forecasting period of 12 observations: 70% of the models passed that forecasting test at 5%, 81% at the 1% level. The best forecasted series belong to the international trade groups, those less satisfactory to the industrial production group. This last result is obviously related to the nonlinearities found in the production series and to the large proportion of outliers occuring in this sector.

References

Bell, W.R. (1995), "X12-Time Series Modeling Course", Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census.

Box, G.E.P and Jenkins, G.M (1976): *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Brockwell, P.J. and Davis, R.A. (1987): *Time Series: Theory and Methods*, New-York: Springer-Verlag.

Chang, I. Tiao, G. and Chen, C., (1988), "Estimation of Time Series Parameters in the Presence of Outliers", *Technometrics*, 30, 193-204.

EUROSTAT (1998), *Monthly Panorama of European Industry*, Luxembourg.

EUROSTAT (1997), *Manual Methodology of Short-Term Statistics, Rules and Recommandations*, Luxembourg.

Fildes, R. and Makridakis S., (1995), "The Impact of Empirical Accuracy Studies on Time Series Analysis and Forecasting", *International Statistical Review*, 63, 3, 289-308.

Fuller, W.A. (1996), *Introduction to Statistical Time Series*, 2nd edition, New York: Wiley.

Gomez, V. (1997), "Automatic Model Identification in the Presence of Missing Outliers and Outliers", Ministerio de Economía y Hacienda, mimeo.

Gomez, V. and Maravall, A. (1996), "Programs TRAMO and SEATS: Instructions for the User", Bank of Spain, Working Paper no.9628, Madrid.

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

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Ljung, G. and Box G.E.P. (1978), "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 65, 297-303.

Maravall, A. (1996), "Unobserved Components in Economic Time Series", *Handbook of Applied Econometrics*, Vol. 1.

McLeod, A.I. and Li, W.K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.

Mélard, G. (1984), "A Fast Algorithm for the Exact Maximum Likelihood of Autoregressive-Moving Average Models", *Applied Statistics*, 35,104-114.

Pierce, D.A. (1978), "Seasonal Adjustment when Both Deterministic and Stochastic Seasonality are Present", in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census, 242-269.

Planas (1998) "Linear Signal Extraction with Intervention Techniques in Non Linear Time Series", EUROSTAT working paper 9/1997/A/1. Forthcoming in *Journal of Forecasting.*

Tiao, G. and Tsay, R. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models", *The Annals of Statistics*, 11, 856-871.

Tsay, R. (1984) "Regression Models with Time Series Errors", *Journal of the American Statistical Association*, 79, 118-124.

Tsay, R. (1986) "Time Series Models Specification in the Presence of Outliers", *Journal of the American Statistical Association*, 81, 132-141.

Table 1 Calendar Effects

Indicators	Production*	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Trading Days	.19	.52	.35	.37	.32	.35
Easter Effect	.12	.23	.21	.13	.09	.14

* The production index should be provided to EUROSTAT adjusted for trading days; however, for some Member States, this is only partly done.

Table 2 Outliers

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Additive Outliers by Series						
no AO	.47	.60	.53	.69	.66	.61
1 AO	.27	.26	.27	.24	.24	.25
> 1 AO	.26	.14	.20	.08	.09	.14
Temporary Changes by Series						
no TC	.73	.78	.76	.81	.80	.78
1 TC	.19	.18	.19	.16	.16	.17
> 1 TC	.08	.05	.05	.03	.03	.05
Level Shifts by Series						
no LS	.75	.70	.80	.73	.78	.75
1 LS	.19	.22	.15	.22	.17	.19
> 1 LS	.05	.07	.04	.05	.05	.05
Outliers by Series						
no Outlier	.32	.37	.37	.43	.44	.39
1 Outlier	.25	.30	.27	.32	.32	.30
> 1 Outliers	.44	.33	.35	.25	.24	.31
Mean # of Outlier by Series						
	1.80	1.34	1.41	.97	1.00	1.25

Table 3 Model Fitted

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Differencing and Stationary Behavior						
Logs.	.78	.88	.82	.83	.88	.84
Stationary Series	.01	.02	.06	.06	.06	.04
Minimum Differencing Required						
Δ	.02	.06	.12	.10	.12	.08
Δ^2	.00	.00	.00	.00	.00	.00
Δ_{12}	.15	.12	.11	.11	.11	.12
$\Delta\Delta_{12}$	.81	.80	.71	.74	.70	.75
Model Specification						
Seasonal Lags	.98	.98	.94	.93	.92	.95
Purely Seasonal	.01	.02	.05	.02	.03	.03
Airline	.65	.60	.56	.62	.60	.61
Ima	.75	.70	.67	.72	.71	.71
Ari	.02	.03	.08	.05	.07	.05

Table 4a Diagnostic Checking

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Uncorrelated Residuals						
Ljung-Box						
at 5%	.92	.90	.91	.95	.96	.93
at 1%	.98	.98	.98	.99	.99	.99
Box-Pierce						
at 5%	.94	.95	.96	.97	.97	.96
at 1%	.99	.99	.99	1.00	1.00	.99
Normal Residuals						
Skewness						
at 5%	.87	.90	.89	.93	.93	.91
at 1%	.96	.98	.98	.99	.99	.98
Kurtosis						
at 5%	.76	.88	.91	.95	.95	.90
at 1%	.86	.94	.97	.99	.99	.95
Independent Residuals						
Ljung-Box on Squared Residuals						
at 5%	.78	.84	.88	.90	.90	.87
at 1%	.87	.92	.95	.96	.96	.93
Box-Pierce on Squared Residuals						
at 5%	.82	.89	.91	.95	.95	.91
at 1%	.89	.95	.97	.98	.98	.96

Table 4b Diagnostic Checking

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Not any significant statistic						
at 5%	.50	.60	.64	.72	.73	.65
at 1%	.72	.82	.87	.91	.92	.86
Models yielding uncorrelated residuals						
at 5%	.88	.87	.88	.92	.93	.90
at 1%	.97	.97	.98	.99	.99	.98
... and normally distributed						
at 5%	.63	.73	.74	.83	.84	.77
at 1%	.83	.91	.93	.97	.97	.93
Evidence of nonlinear dependencies in residuals						
at 5%	.30	.21	.18	.14	.13	.18
at 1%	.19	.12	.08	.06	.05	.09
... related to seasonal behavior						
at 5%	.08	.06	.06	.04	.03	.05
at 1%	.05	.04	.02	.02	.01	.03

Table 5 Models Performances

Acceptable Models by Series

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Airline						
Not any significant statistics						
at 5%	.34	.37	.38	.46	.46	.41
at 1%	.48	.50	.50	.56	.56	.53
... letting uncorrelated residuals						
at 5%	.58	.55	.50	.57	.57	.56
at 1%	.64	.60	.55	.61	.60	.60
Ima						
Not any significant statistics						
at 5%	.39	.43	.44	.53	.53	.48
at 1%	.55	.58	.60	.66	.66	.62
Ari						
Not any significant statistics						
at 5%	.01	.02	.05	.03	.05	.03
at 1%	.01	.02	.06	.04	.06	.04

Table 6 Post-Sample Predictive Tests

Indicators	Production	Turnover	New-Orders	Import	Export	All
# of series	2512	2206	1641	3547	3332	13238
Proportion of satisfying forecasts						
at 5%	.66	.69	.66	.72	.73	.70
at 1%	.76	.81	.77	.84	.83	.81