Journal of Official Statistics, Vol. 16, No. 2, 2000, pp. 173–184

Large Scale Fitting of Regression Models with ARIMA Errors

Björn Fischer[1] *and Christophe Planas*[2]

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 13,238 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 occurrences 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 misspecifications found is detailed, yielding for example an indication on nonlinear patterns actually encountered in monthly series.

Key words: ARIMA models; outliers; diagnostics; seasonal adjustment; forecasting.

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 in respect to some particular features of economic time series. Namely, the trading days rhythm, 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 of the ARIMA-type (see for example Bell 1995; Fuller 1996; Tsay 1984), plus some developments related to outlier detection and correction and to a full automised model identification procedure (see Gomez 1997). These last two advances were crucial for a massive model-based treatment of time series.

[1] Planistat Luxembourg, Building J. F. Kennedy, L-1855 Luxembourg, Luxembourg.
[2] Joint Research Centre, Institute for Systems, Informatics and Safety, TP316, Via E. Fermi, 1, I-21020, Ispra, Italy. Correspondence is to be sent to Christophe Planas. E-mail: christophe.planas@jrc.it.
Acknowledgments: This article is a revised version of the EUROSTAT Working Paper 9/1998/A/8, written on request of EUROSTAT within the framework of a study on Seasonal Adjustment methods. The ideas expressed here are the authors' and do not necessarily reflect the positions of EUROSTAT and of JRC. The authors thank Berthold Feldmann, EUROSTAT-D1, for his interest in this work, Raoul Depoutot, EUROSTAT-A4, and Stefano Nardelli for useful comments. The Associate Editor and two referees are gratefully acknowledged for constructive comments which have substantially improved the presentation of the article.

The quality of the published information 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 13,238 monthly series. Besides the overall capacity of these models in describing our set of 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 occurrences 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 misspecifications found is detailed, yielding for example an indication of nonlinear patterns actually encountered in monthly series.

The sake of rigor leads to underline the limitations of this study. There are three main sources of limitation: the data, the software used, and the selected diagnostics. Our data are monthly and they are set as indexes. Some of them may be correlated; we checked however that cross-correlations could not have an impact on the results. We used the software TRAMO for automatic model identification. Other software may have identified other models. Our choice was based on four main considerations: (1) experiments conducted in EUROSTAT showed that TRAMO is relatively fast (see Dossé and Planas 1996); (2) it is able to identify a wider range of models than an alternative like X11-ARIMA or its successor X12 (see Dagum 1988; Findley et al. 1998); (3) TRAMO is also able to automatically select log-transformations and significant calendar effects; (4) it is routinely used in several units of EUROSTAT. The next limitation is related to the diagnostic checks. We chose a set of tests which covers a general enough range of pathologies. As there is no unanimity about which diagnostics are to be used for automatic checks, other tests could have been used and probably different results would have been obtained. All these points limit the generality of our study.

Nevertheless, as far as we know, no such large scale analysis is available in the literature. While the available computing power has greatly facilitated the use of large databases, we believe that much remains to be learned from the application of statistical techniques to a massive set of series. Accumulation of empirical evidences is useful for the evaluation and improvement of statistical procedures. Our study is a contribution in that direction.

We describe in Section 2 the data collection process and in Section 3 we review the methodological treatment. Section 4 presents the results of the experiment which are further discussed in Section 5.

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. The industrial activities described are sorted according to the revised Classification of Economic Activities within the European Communities (Nace Rev. 1). This classification covers almost all sectors of industrial production in the Member States.

Five different areas are covered: industrial production, turnover, new-orders, imports and exports. The number of series in every group is 2,512, 2,206, 1,641, 3,547, 3,332, respectively, for a total number of 13,238 series. All the series are monthly indexes. The sample sizes are roughly distributed as follows: 10,100 series are of length in [85, 105], 1,500 in [130, 155], 1,300 in [195, 212], the other sample lengths being roughly uniformly distributed outside these intervals between the minimum of 75 and the maximum of 212.

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 occurred in this period. They were split in the same way as 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. Complementary information can be found in EUROSTAT, 1997.

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 six variables plus a length-of-month index, while an Easter effect variable has been specified so as to describe an effect lasting eight 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 1980; Hannan and Rissanen, 1982). The search puts the emphasis on low order and on balanced models; the model with maximum order that can be considered is $(3, 2, 3)\ (1, 1, 1)_{12}$, and the total number of models that can be selected is 384. 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, Tiao and Chen (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 of 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 has been able to adequately describe the properties of the series, an analysis of the residuals has been performed. The statistics used 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; Pierce 1978, respectively). Independency is verified by computing these statistics on the squared residuals (McLeod and Li 1983). As illustrated in Fiorentini and Maravall (1996), computing Q_s on the squared residuals allows to associate nonlinearities with the series seasonal behavior. 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 post-sample predictive testing. The test used is an F-test constructed as the ratio of the mean squared forecast errors to the mean squared residuals.

The test 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 with 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 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 13,238 series of our sample, 35% are sensitive to the trading days rhythm and 14% to Easter recess. The results concerning production indexes are, however, somewhat misleading. Normally, National Statistical Institutes provide EUROSTAT with trading day adjusted production indexes as most of the economic sectors are sensitive to the trading day rhythm. But some series may not have been subject to that treatment, or not in a satisfactory way, and consequently a significant effect was found in 19% of the production series. Among the four other groups, turnover series are the most affected by the trading days with an incidence found in 50% of the

Table 1. Calendar effects

Series	Production[a]	Turnover	New-Orders	Imports	Exports	All
# of series	2,512	2,206	1,641	3,547	3,332	13,238
% of total	19.0	16.7	12.4	26.8	25.2	
Trading Days	.19	.52	.35	.37	.32	.35
Easter Effect	.12	.23	.21	.13	.09	.14

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

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 rhythm.

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, which is quite low. Production is the sector with the highest number of detected outliers (1.80 by series), while import series are those presenting the least data irregularities (.97 by series). For every group, Ao is the most often found outlier-type: 40% of series have at least 1 Ao, against 25% for Ls and slightly less for Tc. More than half of the production series have one or more Ao's. The proportion of Ls and Tc is roughly similar among the different groups.

We now examine the ARIMA specifications that have been identified. Table 3 gives mean results, again displayed by group of series, First, it can be seen that 78% to 88% of the series need a prior log-transformation. For the groups where the indexes are measured 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 than 70% of the cases, nonstationary 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 for very few cases (8% and 12%, respectively). Notice that two regular differences are nearly never needed. Seasonal unit roots are present in more than 80% of the fitted models. Not much discrepancy between groups can be seen with respect to the stationarity properties.

Table 2. Proportion of series with outliers

Series	Production	Turnover	New-Orders	Imports	Exports	All
Additive outliers						
None	.47	.60	.53	.69	.66	.61
One	.27	.26	.27	.24	.24	.25
More than one	.26	.14	.20	.08	.09	.14
Temporary changes						
None	.73	.78	.76	.81	.80	.78
One	.19	.18	.19	.16	.16	.17
More than one	.08	.05	.05	.03	.03	.05
Level shifts						
None	.75	.70	.80	.73	.78	.75
One	.19	.22	.15	.22	.17	.19
More than one	.05	.07	.04	.05	.05	.05
All types						
None	.32	.37	.37	.43	.44	.39
One	.25	.30	.27	.32	.32	.30
More than one	.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

Series	Production	Turnover	New-Orders	Imports	Exports	All
Log Transf.	.78	.88	.82	.83	.88	.84
Amount of differencing selected						
None (statonary)	.01	.02	.06	.06	.06	.04
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
ARIMA form						
Non seasonal	.02	.02	.06	.07	.08	.05
Purely seasonal	.01	.02	.05	.02	.03	.03
Multiplicative	.97	.96	.89	.90	.89	.92
Stationary	.01	.02	.06	.06	.06	.04
Airline	.65	.60	.56	.62	.60	.61
Other IMA	.10	.09	.11	.09	.10	.10
ARI	.01	.02	.04	.01	.02	.02
Mixed ARIMA	.23	.27	.24	.22	.21	.23

Regarding the ARIMA specification, in Table 3 a first distinction is made between non seasonal, purely seasonal, and multiplicative models. It is seen that purely seasonal or purely regular models were used in less than 5% of the cases, respectively, against 90% for multiplicative models. This result underlines the importance of multiplicative models for describing the behavior of monthly economic indicators. Models are then classified as stationary, airline, other IMA, ARI, and mixed ARIMA models. The airline model, introduced by Box and Jenkins (1976), is the most simple multiplicative IMA; it is specified as $\Delta\Delta_s x_t = (1 + \theta_1 B)(1 + \theta_s B^s)a_t$, where s denotes the data-periodicity, B the lag operator, Δ and Δ_s are differencing operators at lags 1 and s, and a_t are the innovations. This model has been widely used both in applied works and in methodological studies.

Table 3 shows that TRAMO selects the airline model in 60% of the cases, against 10% for higher order IMA, 4% for pure ARI, and 23% for mixed forms. The remaining cases are made up of stationary ARMA's. Notice that as the model identification procedure eventually relies on the BIC, a model is always proposed. The model adequacy must, however, be verified.

Table 4a displays the diagnostic checks on the fitted models. 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 shows less than 4% departures from the white noise hypothesis, which is less than the nominal size. Regarding the residual distribution, only 10% of the overall set of residuals show a significant departure from normality. Yet, this proportion reaches 24% for residual kurtosis in industrial production indexes. It is the sector where most 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

Table 4a. Diagnostic checking

Series	Production	Turnover	New-Orders	Imports	Exports	All
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

the residual distribution, tests for residual independency show that roughly 10% of the series embody some significant nonlinear structure. The group most affected seems to be industrial production. All the departures found are not strongly evident: considering 1% nominal sizes for the tests reduces the proportion of rejection to 1% for correlation and normality statistics, and to 7% for linear independency tests.

Table 4b further summarizes the diagnostics. For 65% of the series, the modeling is entirely adequate, in that no 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 fully adequate in only 50% of the cases. Lowering at 1% the significance level of every test increases the number of acceptable models from 72% to 92% for these two groups which remain the extremes, and to 86% for the overall set of series. This is a positive result about the capacity of ARIMA models with exogenous regressors to describe economic time series. Furthermore, if interest focuses on models able to describe second moments of the series, then it is seen in Table 4b that 90% of the fitted models let residuals with white noise properties, 98% at the 1% level.

The departures from correct model specification are also of interest. The distribution of 23% of the residuals presents 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

Table 4b. Diagnostic checking

Series	Production	Turnover	New-Orders	Imports	Exports	All
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

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 raised the number of satisfying models found for these indicators.

It is interesting to evaluate the performances of the different model specifications in fitting the data. For every type of model, Table 5 reports the proportion of fits satisfying all diagnostic checks. The number of times a model specification has been used is also reported. The proportion of satisfactory fits is roughly similar for the different specifications, around 65% and 85% at the 5% and 1% levels, respectively. The airline model, which was the most used, yields slightly better results, while on the other hand, mixed models are slightly less satisfying. In more than 90% of the cases, the airline model yields uncorrelated residuals. The scarce use of ARI models is more related to the automatic model procedure of TRAMO which favors balanced and low-order models rather than the potential performances of autoregressive models.

Finally, Table 6 shows the results of post-sample predictive tests (see Harvey 1989, p.271). 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 this direction here. Not only would this be outside our scope, but the methodology we have used which also performs automatic corrections for both outliers and calendar effects is more general than 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 forecasting period of 12 observations: 70% of the models passed that forecasting test at 5%, 81% at the 1% level.

Table 5. Model performances

Series	Production	Turnover	New-Orders	Imports	Exports	All
Airline						
Used	1,625	1,334	911	2,182	2,005	8,057
Not any significant statistics						
at 5%	.52	.62	.68	.74	.76	.67
at 1%	.75	.83	.90	.92	.93	.87
... letting uncorrelated residuals						
at 5%	.90	.91	.92	.94	.95	.93
at 1%	.99	.99	.99	.99	1.00	.99
Other IMA						
Used	262	198	179	335	338	1,312
Not any significant statistics						
at 5%	.44	.64	.54	.74	.71	.63
at 1%	.63	.83	.89	.92	.92	.84
ARI						
Used	26	41	61	39	76	243
Not any significant statistics						
at 5%	.54	.49	.66	.64	.67	.62
at 1%	.77	.78	.82	.87	.88	.84
Mixed ARIMA Model						
Used	568	587	398	790	702	3,045
Not any significant statistics						
at 5%	.45	.56	.59	.66	.68	.60
at 1%	.69	.78	.82	.90	.90	.83

The best forecasted series belong to the international trade groups, the 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.

5. Discussion

Following the suggestion of a referee, we checked the differences between the Member States. Focusing on a selection of seven countries, namely D, E, F, I, NL, SE, U.K., we

Table 6. Post-sample predictive tests

Series	Production	Turnover	New-Orders	Imports	Exports	All
Proportion of satisfying forecasts						
at 5%	.66	.69	.66	.72	.73	.70
at 1%	.76	.81	.77	.84	.83	.81

found that in general the results are sufficiently homogeneous between these countries. Some interesting discrepancies could however be seen. The nonlinearities in the production series are mainly due to the German and Spanish production indexes: at the 5% level, 51% of the German production series and 47% of the Spanish ones yielded residuals which presented evidences of nonlinear dependencies. For the production indexes in these two countries, at the 5% level only 23% of the fits satisfied all our diagnostic checks. For Spain, this result is associated with an average number of outliers by series relatively high, at 2.83 against an average of 1.80 by series over all countries. These outliers are mainly of the Ao type: 72% of Spanish production indexes embody one or more outliers, against 53% for all countries. For German production series, the nonlinearities seem to be related to some heteroskedastic pattern, since only 51% of the series pass the post-sample predictive test against 66% in mean over all countries. We let the problem of explaining these nonlinear behaviors be an open issue for applied economists.

We found other discrepancies about the sensitivity of the series to trading days. For U.K. turnover series, a significative trading day component could be found in 75% of the series, against 52% on average over all countries. French import and export series embody that effect in 67% and 53% of the series, against respectively 37% and 32% in the mean. A last noticeable fact regards Netherland's new-orders and international trade series which were found to yield slightly more outliers.

In this study, we found many models yielding diagnostics acceptable at the 1% level, but not at the 5% level (see Tables 4a-4b and 6). These are cases where the statistical evidence was not particularly strong. As pointed out by a referee, it is possible that in these cases, supplementing automatic procedures with expert modelling could make the results acceptable at the 5% level. Improvements might be obtained by focusing for example on the ARMA model specification or on the outlier treatment.

In conclusion, the result of this large scale experiment is encouraging: time series regression with ARIMA errors seems to be a powerful tool for the automatic analysis of massive sets of monthly series. We find most striking the result that in 90% of the cases, 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 about the proportion of models passing all diagnostic together with 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).

Given the diagnostic checks considered, the ability of TRAMO's automatic model identification procedure to yield well-specified models is remarkable. For a production concern, TRAMO gives access to a reliable automatic model-based treatment at low cost. However, when few series are considered, we believe that the choice of a model-based approach should only be seen as a starting point: in particular, a good fit does not necessarily imply a good seasonal adjustment. An important requirement for performing a reliable seasonal adjustment is the knowledge of the series under analysis. Exhaustive prior knowledge allows plausibility checks, for example about breaks, calendar effects or regression parameters, which are useful complements of any automatic method. Outliers, especially at the current end of the time series, need to be linked with background information, so that it is possible to study their sources and to find an appropriate way to handle them. Furthermore, the study of the outlier empirical distribution offers the opportunity to detect important regression effects that had not been explicitly modelled, like moving school holidays, moving dates of fairs, and also inconvenient datatransformations. Altogether, we believe that a careful use and control of regression models with ARIMA errors offers to practitioners a powerful tool for improving the quality of seasonally adjusted figures.

6. References

Bell, W.R. (1995). X12-Time Series Modeling Course. Washington, D.C.: U.S. Department of Commerce, U.S. 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.

Dagum, E.B. (1988). The X11-ARIMA/88 Seasonal Adjustment Method - Foundations and User's Manual. Time Series Research and Analysis Division, Statistics Canada, Ottawa.

Dossé, J. and Planas, C. (1996). ''Pre-Adjustments in Seasonal Adjustment: A Comparison of REGARIMA and TRAMO''. EUROSTAT Working Paper no. D3/SA/07. Luxembourg.

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

EUROSTAT (1997). Manual Methodology of Short-Term Statistics, Rules and Recommendations, 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.

Findley, D., Monsell, B., Bell, W., Otto, M., and Chen, B.C. (1998). New Capabilities and Methods of the X12-ARIMA Seasonal Adjustment Program. Journal of Business and Economic Statistics, 16, 2, 127–152.

Fiorentini, G. and Maravall, A. (1996). Unobserved Components in ARCH Models: An Application to Seasonal Adjustment. Journal of Forecasting, 15, 175–201.

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 Econimía 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. (1980). The Estimation of the Order of ARMA Processes. Annals of Statistics, 8, 1071–1081.

Hannan, E.J. and Rissanen, J. (1982). Recursive Estimation of Mixed Autoregressive-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.

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. Department of Commerce, U.S. Bureau of the Census, 242–269.

Planas (1998). Linear Signal Extraction with Intervention Techniques in Non Linear Time Series. Journal of Forecasting, 17, 515–526.

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.

Received May 1998
Revised January 2000