To appear (with minor modifications) as Chapter 7 in A Course in Time Series Analysis, D. Peña, G.C. Tiao, and R.S. Tsay (eds.) NY: J. Wiley and Sons (expected date: June 2000)

Automatic Modeling Methods for Univariate Series

Víctor Gómez and Agustín Maravall

November 1998

Chapter

Automatic Modeling Methods for Univariate Series

In this chapter, a unified approach to automatic modeling for univariate series is presented. First, ARIMA models and the classical methods for fitting these models to a given time series are briefly reviewed. Second, some automatic methods for model identification are described and an algorithm for automatic model identification is proposed. Third, outliers are incorporated into the model and an algorithm for automatic outlier detection and correction is proposed. Fourth, combining the proposed algorithms for automatic model identification and automatic outlier detection and correction, an algorithm is proposed for automatic model identification in the presence of outliers. Finally, the previous algorithm is extended to cope with missing observations, trading day and Easter effects, and intervention and regression effects.

1.1 Classical Model Identification Methods

In this section, we briefly review the classical model identification methods, with particular emphasis in the Box and Jenkins' method. It is pointed out that the classical methods are subjective, in the sense that the results depend to a great degree on the analyst's experience and background. In addition, it is noted that the tools proposed by Box and Jenkins for model identification work well for pure moving average or pure autoregressive models, but not so well for mixed ARMA models.

It has already been mentioned in previous chapters that ARIMA models can be successfully used in practice to represent many time series, especially

in economics. The general seasonal multiplicative ARIMA model is

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D z_t = C + \theta(B)\Theta(B^s)a_t, \tag{1.1}$$

where, C is a constant, s is the number of seasons, $d = 0, 1, 2$, $D = 0, 1$, $\nabla = 1 - B$ is a regular difference, $\nabla_s = 1 - B^s$ is a seasonal difference, and B is the backshift operator, $Bz_t = z_{t-1}$. Instead of z_t, it may be necessary to use $log(z_t)$, or some other transformation, to stabilize the variance of the series. The roots of the autoregressive polynomials, $\phi(B)$ and $\Phi(B)$, are assumed to lie outside and those of the moving average polynomials, $\theta(B)$ and $\Theta(B)$, on or outside of the unit circle. If p and P are the degrees of $\phi(B)$ and $\Phi(B)$ and q and Q those of $\theta(B)$ and $\Theta(B)$, model (1.1) is denoted as a multiplicative $(p, d, q)(P, D, Q)_s$ model.

ARIMA models became very popular after the publication of the seminal book by Box and Jenkins (1976), whose first edition was published in 1970. In this book, a systematic procedure was for the first time proposed to handle the problem of time series modeling by means of iterative cycles consisting of

i) Model identification

ii) Model estimation

iii) Diagnostic checking

At the same time, in Box and Jenkins' book of 1976 some computer programs were given to implement this methodology with the aid of digital computers. In this way, given the big computing power of computers, the user was for the first time able to apply a new and powerful machinery which would prove very useful in the field of time series analysis. The arrival of personal computers contributed further to increase enormously the demand for these new tools.

At the identification stage one selects a particular model from the class (1.1), that is, one selects values for p, d, q, P, D and Q. Classical tools for model identification include plotting the data over time, the sample autocorrelation function, the sample inverse autocorrelation function and the sample partial autocorrelation function. All these functions have been defined in previous chapters. Also, steps ii) and iii) have been fully described in chapters 2, 3, and 4, so our focus in this chapter will be on step i).

The modeling procedure proposed by Box and Jenkins (1976) was by no means a process that could be fully automated with the help of computers.

In particular, step i) of the procedure was rather an art which required an expert in time series analysis to carry it out. It was certainly the most difficult of the three steps. In spite of the advances that have taken place in the last two decades, we can say that step i) continues to be the most difficult of the three steps. Besides, we have to be aware of the fact that the majority of time series encountered in practice usually have outliers, which makes even more difficult the modeling procedure.

We will briefly review later the most important methods for automatic detection and correction of outliers that are currently in use. As far as ARIMA model identification is concerned, the presence of outliers can make it very difficult due to the important biases induced by the outliers in the parameter estimates and in the sample autocorrelation and partial autocorrelation functions. For this reason, any good strategy for ARIMA model identification has to account for the presence of outliers.

There are many reasons why one should try to automate as much as possible the ARIMA model identification stage, but they can be basically reduced to two. The first one is that one should eliminate as much as possible all mundane and mechanical chores, which can be performed by the computer, thus increasing the analyst's productivity. If the user is an accomplished analyst, he may invest more of his precious time on troublesome data sets that he has to model. On the contrary, if he is not an expert in time series models, he can use a powerful methodology that he couldn't even dream of using before. The second reason has to do with the objectivity of the identification stage, since it is desirable that this stage be not subject to heuristic methods and ad-hoc procedures that vary with each time series expert. For example, if a National Statitical Office has to produce some statistical data which require the modeling of some time series datasets and an expert is involved in the production process who uses subjective techniques, it may be criticized for publishing data which are neither objective nor reproducible.

1.1.1 Subjectivity of the Classical Methods

The Box and Jenkins' method for model identification relies heavily on the inspection of plots of data over time and the inspection of the graphs of the sample autocorrelation and partial autocorrelation functions. These last tools can be effective to identify pure autoregressive or pure moving average models, but no so effective with mixed ARMA models. Besides, the determination of the stationary transformation, that is, the numbers d and D in (1.1), can be very difficult.

With the exception of very few cases in which the data show a very distinctive pattern, it is usually rather difficult to identify a model for the series at hand. For example, given a sample of finite length, it may be extremely difficult to distinguish between the nonstationary model $(1 - .7B)\nabla z_t = a_t$ and the process $(1 - 1.704B + .706B^2)z_t = (1 - .715B)(1 - .989B)a_t$, which has very similar coefficients but for which the autoregressive polynomial has all its zeros outside the unit circle.

Therefore, it is most probable that several time series experts who use the Box and Jenkins' method, when confronted with the same data, will specify different models. This makes the whole process of classical model identification dependent on the person who applies the techniques. The more experience he has, the more likely he is to select an adequate model for the series. This subjectivity is inherent in the classical model identification methods.

1.1.2 The Difficulties With Mixed Arma Models

It has already been mentioned in the last section that the sample autocorrelation and partial autocorrelation functions can effectively identify pure moving average models and pure autoregressive. On the other hand, when both the degrees of the autoregressive polynomial $(p + P)$ and the moving average polynomial $(q + Q)$ are not 0, the previous functions are much more difficult to interpret. In this case, other model identification methods, different from the classical methods, are called for.

The difficulty of identifying mixed ARMA models is further increased when seasonality is also present in the time series at hand. Several major advances have been made in the last two decades to identify ARIMA models for non-seasonal time series. Among these, we can mention the extended autocorrelation function and the smallest canonical correlation methods developed by Tsay and Tiao (1984, 1985). These methods are very informative in the identification of ARIMA models for non-seasonal time series, but they are less successful when they are directly applied to seasonal time series. It is to be noted that these methods can also be used with nonstationary series.

Since the early 1970s, some penalty function methods have been proposed for ARMA model identification. These methods can be used with seasonal time series and their popularity is constantly increasing. The reason for this is that they are automatic and can be effective and computationally cheap. However, although some results have been extended to nonstationary series, these methods are in principle only applicable to stationary series.

1.2 Automatic Model Identification Methods

In this section, we deal with automatic model identification methods, in contrast to the classical model identification methods considered in the last section. First, in order to obtain the degrees d and D of the stationary transformation in (1.1), we can use unit root tests. Then, several methods can be applied to identify an ARMA model for the stationary (differenced) series. We review in this section the penalty function and the pattern identification methods. Both of these methods are automatic and can be regarded as objective. However, there is always some degree of subjectivity also in these methods, e.g. when selecting the highest orders, p, P, q and Q,to be considered for model (1.1). For this reason, we prefer to use the term automatic rather than objective when we refer to them. A good reference for ARMA model identification is the book by Choi (1992).

1.2.1 Unit Root Testing

In the Box and Jenkins methodology, the decision concerning the need for differencing is based upon the characteristics of the plot of the data and of its sample autocorrelation function. For example, failure of this last function to die out sufficiently quickly indicates that differencing is required.

In recent times, there has been a growing interest in more formal inference procedures concerning the appropriateness of differencing operators in the model. Since all the roots of the differencing operators $\nabla = 1 - B$ and $\nabla_s = 1 - B^s$ lie on the unit circle, testing for differencing is usually referred to as unit root testing.

It is interesting to note that, as Dickey and Pantula (1987) point out, the results obtained by several authors suggest that overdifferencing is not a problem as far as forecasting is concerned. However, there appear to be uses for unit root tests in investigating some economic hypothesis. The practical implication of this is that when one is interested in the routine treatment of many series for forecasting purposes, one should not care very much about whether or not some of the series are overdifferenced. It is our practical experience that much the same thing happens with regard to model based seasonal adjustment. We can say that overdifferencing is compensated by moving average parameters that go to unity.

We will not review here the vast amount of existing literature concerning unit root testing. The reader can consult, for example, Reinsel (1997) and the references therein. We will content ourselves with making a few remarks

on existing procedures.

The two "classical" unit root tests of Dickey–Fuller and Phillips–Perron tend to exhibit rather poor behavior in the presence of certain types of serial correlation. See the Monte Carlo analysis of Schwert (1989).

When there is no seasonality in the series at hand and only regular differences, that is, differences of the form ∇^d, are considered, it seems that the sequential testing procedure suggested by Dickey and Pantula (1987) is the best strategy to follow. According to these authors, only tests that compare a null hypothesis of k unit roots with an alternative of $k-1$ unit roots are considered. In the sequential procedure, one should start with the largest k under consideration and work down, that is, decrease k by one each time the null hypothesis is rejected.

The situation is different for seasonal time series. In this case, further research is needed and no general agreement exists on how to proceed as far as unit root testing is concerned. It seems that a generalization of the Dickey and Pantula (1987) approach to the seasonal case would be an interesting topic to investigate.

1.2.2 Penalty Function Methods

In the identification stage, once the differencing orders d and D in (1.1) have been obtained for the nonstationary series $\{z_t\}$, the problem remains of finding an ARMA model for the differenced series $u_t = \nabla^d \nabla_s^D z_t$.

Since the early 1970s, some procedures to determine the orders k and i of an ARMA(k,i) model have been proposed which minimize a function of the form

$$P(k,i) = \ln \hat{\sigma}^2_{k,i} + (k+i)\frac{C(n)}{n}, \qquad k \leq K, \qquad i \leq I, \tag{1.2}$$

where $\hat{\sigma}^2_{k,i}$ is the maximum likelihood estimate of the variance of the white noise variance, $C(n)$ is some function of the number of observations n of the series, and K and I are upper bounds for the orders, usually imposed a priori. Because $\hat{\sigma}^2_{k,i}$ decreases as the orders increase, it cannot be a good criterion to select the orders by minimizing it. This is the reason why the penalty term $(k+i)C(n)/n$ is included.

If $C(n)$ in (1.2) is replaced with 2, we obtain the famous AIC criterion, which stands for Akaike's Information Criterion. Other possible choices are $C(n) = \ln n$, which corresponds to the BIC (Bayesian Information Criterion), and $C(n) = 2\ln(\ln n)$, which gives the HQ criterion (Hannan and Quinn). The BIC criterion imposes a greater penalty term than does AIC.

One criterion for selection of AR(p) models is the FPE (Final Prediction Error) criterion, which is given by $\mathrm{FPE}(p) = \{1 + (p/n)\}\hat{\sigma}_p^2$.

The BIC criterion estimates the orders of an ARMA model consistently, whereas the AIC does not. However, this is not a reason to prefer BIC instead of AIC because consistency is based on the assumption that there is a "true" ARMA model for the series and this is doubtful proposition. Models are artificial constructs and probably there is no such a thing as a true model.

The FPE, AIC and BIC criteria have been described in more detail in chapter 5.

It is our practical experience and also the experience of some other authors, like, for example, Lütkepohl (1985), that the BIC criterion works better in practice than AIC, in terms of selecting more often the original model when working with simulated series and selecting models with a better fit when working with real series.

Although the penalty function methods are in principle computationally expensive, because they need maximum likelihood estimates for all possible ARMA models, there are methods, like the Hannan-Rissanen's method described later in this chapter, which use cheaper estimates based on linear regression techniques only. Also, in the case of multiplicative seasonal ARMA models, it will be seen that it is possible to further reduce the computational burden by proceeding sequentially. That is, by iterating between selections of the regular and of the seasonal parts.

The penalty function methods can also be used to identify vector ARMA models, see Reinsel, (1997). The penalty functions to use with multivariate data are direct generalizations of the ones for the univariate case. This is a great advantage, not shared by many of the other identification methods.

1.2.3 Pattern Identification Methods

Since the early 1980s, some methods have been applied for determining the orders of an ARMA process which use the extended Yule Walker equations. For the ARMA(p, q) process

$$z_t + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} = C + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

these last equations are given by

$$\gamma_j = -\phi_1 \gamma_{j-1} - \cdots - \phi_p \gamma_{j-p}, \qquad j = q+1, q+2, \ldots,$$

where γ_j, $j = 0, 1, \ldots$, is the autocovariance function of the process. These methods are often called pattern identification methods. See Choi (1992).

It is to be noted that, contrary to Choi's remark about penalty function methods being computationally exorbitant and pattern identification methods being computationally cheap, it will be shown later in this chapter that the proposed sequential application of Hannan-Rissanen's method, which is based on the BIC criterion, for stationary seasonal models is computationally cheap and can be very effective.

The pattern identification methods are so called because they are based on certain functions which give rise to two-way arrays with distinctive patterns. For each ARMA(p, q) model, the corresponding two-way array shows a unique pattern. Using the sample analog of this two-way array, an ARMA model is identified by looking for a theoretical pattern which is closely resembled by the sample one. Among the many pattern identification methods which have been proposed in the literature, we can mention the R and S array method by Gray, Kelley and McIntire (1978), the Corner method by Beguin, Gourieroux and Monfort (1980), the extended sample autocorrelation method by Tsay and Tiao (1984) and the smallest canonical correlation method by Tsay and Tiao (1985). These last two methods can be effective with non-seasonal time series and can also be used with nonstationary series. However, the R and S array method and the Corner method, which can only be used with stationary series, do not seem to be very useful even for data with no seasonality. The Corner method has been applied to identification of transfer function models by Liu and Hanssens (1982).

1.2.4 Uniqueness of the Solution and the Purpose of Modeling

In the identification stage of model building, it is often the case that there are several models for which the fit is acceptable. For example, if the BIC criterion is used, there may be a very small difference between the BIC of an AR(2) model and the BIC of an ARMA$(1, 1)$ model. In this case, we can probably use any of these two competitive models to model the data.

When some competitive models exist, one should try to select the more parsimonious one. That is, the one with less parameters. On occasion, it may be useful to select models that are also balanced. This means that the degree of the autoregressive part, included the nonstationary transformation, equals the degree of the moving average part. Balanced models are useful when one is going to perform model based seasonal adjustment.

In summary, models should be considered as artificial constructs, which are useful for certain purposes, but are only a crude approximations to rea-

lity. In this respect, the criteria used to select models, especially when some competing models exist, depend on the applications. Some criteria may be good for forecasting, but not so good for signal extraction, for example. One should always have in mind that usually there is not a unique solution to the identification problem.

1.3 Tools for Automatic Model Identification

In this section, some practical procedures will be described for automatic model identification. The emphasis is in the word practical, so that the methods presented will aim at simplicity, efficiency and speed when applied to real data.

We will start with a test which we propose for the log-level specification. The test is based on the maximum likelihood principle applied to a series which is supposed to follow the model $(0,1,1)(0,1,1)_s$. This is the airline model of Box and Jenkins (1976). The reason why we select that model in this and other tests later in this chapter is that it encompasses many other models and is a model very often found in practice.

We will then review the two-stage method proposed by Gómez (1998) to estimate unit roots. After that, the Hannan-Rissanen's method, hereafter referred to as the HR method, will be reviewed. This method is used to identify an ARMA model for the stationary (differenced) series. It is based on the BIC criterion and is computationally cheap. Some improvements to the HR method, proposed by Gómez (1998), will be described.

1.3.1 Test for the Log-Level Specification

The test for the log-level specification is based on the maximum likelihood estimation of the parameter λ in the Box-Cox transformations. We fit an airline model with mean to the data, first in logs ($\lambda = 0$) and then without logs ($\lambda = 1$). Let $z = (z_1, \ldots, z_n)'$ be the differenced series and let T be a transformation of the data, which can be any of the Box-Cox transformations. It is assumed that $T(z)$ is normally distributed with mean 0 and $\mathrm{Var}(T(z)) = \sigma^2\Sigma$. Then, the logarithm of the density function $f(z)$ of z is

$$\ln\left(f(z)\right) = k - \frac{1}{2}\left\{n\ln(\sigma^2) + \ln|\Sigma| + T(z)'\Sigma^{-1}T(z)/\sigma^2 + \ln(1/J(T))^2\right\},$$

where k is a constant and $J(T)$ is the jacobian of the transformation. Considering the parameter λ in the T transformation fixed, the previous density

function is maximized first with respect to the other model parameters. It is easy to see that σ^2 can be concentrated out of this function by replacing it with the maximum likelihood estimator $\hat{\sigma}^2 = T(z)'\Sigma^{-1}T(z)/n$. The concentrated function is

$$l(z) = -\frac{1}{2}\left\{n \ln(T(z)'\Sigma^{-1}T(z)) + \ln|\Sigma| + \ln(1/J(T))^2\right\} + \cdots,$$

where the dots indicate terms that do not depend on the model parameters. After having maximized with respect to all model parameters different from λ, we maximize with respect to λ. Let $\Sigma = LL'$, where L is a lower triangular matrix, be the Cholesky decomposition of Σ. Then, the expression $T(z)'\Sigma^{-1}T(z) \times |\Sigma|^{1/n} = |L|^{1/n}T(z)'\Sigma^{-1}T(z)|L|^{1/n}$ is a nonlinear sum of squares which we denote by $S(z,T)$. The maximum likelihood principle leads to the minimization of the quantity $S(z,T)(1/J(T))^{2/n}$. It is easy to see that $(1/J(T))^{1/n}$ is the geometric mean in the case of the logarithmic transformation, and unity in the case of no transformation. Therefore, the test compares the sum of squares of the model without logs with the sum of squares multiplied by the square of the geometric mean in the case of the model in logs. Logs are taken in case this last function is the minimum.

1.3.2 Regression Techniques for Estimating Unit Roots

Let the observed series $\{z_t\}$ follow the ARIMA(p,d,q) model

$$\phi(B)(\delta(B)z_t - \mu) = \theta(B)a_t, \tag{1.3}$$

where $\phi(B) = 1 + \phi_1 B + \ldots + \phi_p B^p$, $\delta(B) = 1 + \delta_1 B + \ldots + \delta_d B^d$ and $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$ are polynomials in the backshift operator B of degrees p, d and q, $\{a_t\}$ is a i.i.d. $N(0,\sigma^2)$ sequence of random variables and μ is the mean of the differenced process. The roots of $\delta(B)$ are assumed to lie on and those of $\phi(B)$ outside the unit circle, so that the process $w_t = \delta(B)z_t$ follows a stationary ARMA(p,q) process. As mentioned earlier, most economic series follow so-called multiplicative seasonal models, where

$$\begin{aligned} \delta(B) &= \nabla^d \nabla_s^D, \\ \phi(B) &= \phi_r(B)\phi_S(B^s), \\ \theta(B) &= \theta_r(B)\theta_S(B^s), \end{aligned} \tag{1.4}$$

s is the number of observations per year, $\nabla^d = (1-B)^d$ and $\nabla_s^b = (1-B^s)^D$. In practice, for economic time series, the inequalities $0 \leq d \leq 2$ and $0 \leq D \leq$

1 hold. For simplicity, we will use in the rest of the section the notation (1.3), even for multiplicative seasonal models. We will make specific reference to these models when necessary.

In the following, a procedure to obtain the differencing orders is reviewed which is based on the estimation of unit roots. This last procedure is the first step of an automatic model identification method which has been proposed by Gómez (1998) and is implemented in programs TRAMO and SEATS, see Gómez and Maravall (1997). The estimation of the unit roots is done by first estimating autoregressive models of the form

$$(1 + \phi_1 B + \phi_2 B^2)(1 + \Phi B^s)(z_t - \mu) = a_t, \tag{1.5}$$

where $\{z_t\}$ is the observed series, s is the number of observations per year, μ in the mean of the process and $\{a_t\}$ is a sequence of i.i.d. $N(0, \sigma^2)$ random variables. Then, the series is differenced using the differencing orders given by the unit roots obtained after estimating (1.5) and an ARMA$(1,1)\times(1,1)_s$ model with mean, that is, a model of the form

$$(1 + \phi B)(1 + \Phi B^s)(x_t - \mu) = (1 + \theta B)(1 + \Theta B^s)a_t, \tag{1.6}$$

is fitted to the differenced series $\{x_t\}$. If any new unit roots appear after estimating (1.6), the differencing orders are properly increased and a new model (1.6) is fitted. The process is continued until no more unit roots are found. Then, the residuals of the last estimated model are used to decide whether to specify a mean for the model or not. The choice of models (1.5) and (1.6) will be justified later.

Suppose that the series $\{z_t\}$ follows model (1.3), where it is assumed $\mu = 0$ to simplify matters. Then, by theorems 3.2 and 4.1 of Tiao and Tsay (1983), the ordinary least squares (OLS) estimators obtained from an AR(k) regression, where $k \geq d$, asymptotically verify

$$\hat{\Phi}_k(B) \doteq \hat{\delta}(B)\hat{\phi}_m(B),$$

where $\doteq$ denotes asymptotic equivalence in probability, $m = k - d$ and $\hat{\Phi}_k(B)$, $\hat{\delta}(B)$ and $\hat{\phi}_m(B)$ are, respectively, the polynomials estimated by OLS in the autoregressions

$$\begin{aligned} \Phi_k(B)z_t &= a_t, \\ \delta(B)z_t &= a_t, \\ \phi_m(B)w_t &= a_t, \end{aligned}$$

where $w_t = \delta(B)z_t$ is a stationary process that follows the ARMA(p,q) model $\phi(B)w_t = \theta(B)a_t$ and the subindex in $\Phi_k(B)$ and $\phi_m(B)$ denotes the polynomial degree. In addition, the equality $\hat{\delta}(B) = \delta(B) + O_p(N^{-1})$ holds, where N is the series length.

The practical implication of this result is that if we perform an autoregression of order greater than or equal to the (unknown) degree of the polynomial $\delta(B)$, we obtain a consistent estimate of $\delta(B)$ as a component of $\hat{\Phi}_k(B)$. If we specify a model of the form AR$(2)\times(1)_s$ for $\Phi_k(B)$, we cover the cases $\delta(B) = 1$, $\delta(B) = \nabla$, $\delta(B) = \nabla_s$, $\delta(B) = \nabla\nabla_s$ and $\delta(B) = \nabla^2\nabla_s$, which are the ones of most applied interest.

In the case of non–seasonal models, where $\delta(B) = \nabla^d$ and $0 \leq d \leq 2$ is assumed, if we specify an AR(2) model, all important cases are covered.

Based on the previous considerations, the algorithm to identify the differencing polynomial is the following:

I) Specify a model of the form AR$(2)\times(1)_s$ with mean, given by equation (1.5) if the process is multiplicative seasonal, or an AR(2) model with mean, also given by (1.5) but without the second factor, if the process is regular. This autoregressive process is estimated using the HR method, which will be described later, unless the user decides to use unconditional least squares. If the roots estimated with the HR method lie outside the unit circle, the autoregression is estimated again using unconditional least squares. A root is considered to be a unit root if its modulus is greater than a specified value, which by default is .97. Go to II).

II) In addition to the differencing degrees identified in I) as a result of the estimated unit roots, a model of the form ARMA$(1,1)\times$ $(1,1)_s$ with mean for seasonal series, or a model ARMA$(1,1)$ with mean for non–seasonal series, is specified. Letting x_t be the series that results from differencing z_t with the differencing polynomial obtained after the estimation of the initial autoregression, the equations for these models are given by (1.6) in the seasonal case, and by (1.6) without the factors involving B^s in the regular case. The model is estimated using the HR method or exact maximum likelihood, depending on the option selected by the user, and if any of the estimated autoregressive parameters is close to 1, the degree of differencing is increased accordingly. A parameter is considered to be close to 1 if its modulus is greater than a specified value, which by default is .88. To avoid cancellation of terms in the model, the absolute value of the difference between each auto-

regressive parameter and its corresponding moving average parameter should be greater than .15. For multiplicative seasonal models, it is not possible to pass from 0 differencing to $\nabla\nabla_s$ directly. If this happens, the roots of the autoregressive polynomial obtained in I) are considered again, the one with greatest modulus is selected, and the series is differenced accordingly. If the series has been differenced in this step, repeat II). Otherwise, go to III).

III) Using the residuals of the last estimated model, it is decided whether to specify a mean for the model of the series or not depending on the significance of the estimated residual mean. Stop.

The ARMA$(1,1)\times(1,1)_s$ model used in II) is very flexible and constitutes a generalization of the airline model of Box and Jenkins (1976). For stationary series, it approximates well many of the ARMA models encountered in practice. When it is used with nonstationary series, it can detect autoregressive unit roots which have not been detected by the autoregressive model used I). Imagine, for example, a model of the form $(1-B)z_t-\mu=(1-.8B)a_t$, where the autoregressive and the moving average part almost cancel out. In this case, an ARMA$(1,1)$ model would probably estimate the unit root better than an AR(2) model.

Consider now the case of a regression model with ARIMA errors. The question naturally arises as to whether the previous analysis is still valid and if, in consequence, the procedure just described is also applicable in this case. By the results of Tsay (1984), pp. 119–120, it is possible, under very general conditions, to work with the original series in order to identify the differencing polynomial.

1.3.3 The Hannan and Rissanen's Method

After having obtained the stationary transformation, the next step in the model building process is the identification of an ARMA(p,q) model for the differenced series, possibly corrected for outliers and other regression effects. We will start by assuming that there are neither outliers nor other regression effects and we will extend the results later in this chapter to the general case.

In the following, the HR method and a procedure to identify ARMA (p,q) models based on it are reviewed. This last procedure is the second step of an automatic model identification method proposed by Gómez (1998) and is implemented in programs TRAMO and SEATS, see Gómez and Maravall (1997). The HR method is a penalty function method based on the BIC

criterion, where the estimates of ARMA model parameters are computed by means of linear regressions. Therefore, these estimates are computationally cheap, although it can be shown that the estimators have similar properties to those obtained by maximum likelihood. See Hannan and Rissanen (1982).

Let $z = (z_1, \ldots, z_n)'$ the observed series, which follows model (1.3), where we assume $\mu = 0$ for simplicity. After $\delta(B)$ has been identified, we can compute the differenced series $w_t = \delta(B)z_t$, $t = d+1, \ldots, n$, which follows the ARMA(p, q) model

$$\phi(B)w_t = \theta(B)a_t, \tag{1.7}$$

where $\phi(B)$, $\theta(B)$ and $\{a_t\}$ are like in (1.3). If the model is multiplicative seasonal, the decomposition (1.4) holds. In order to avoid notational problems, let the differenced series be $w = (w_1, \ldots, w_{n-d})'$. If the orders of the fitted model (1.7) are (p, q), the BIC statistic is

$$\mathrm{BIC}_{p,q} = \log(\hat{\sigma}^2_{p,q}) + (p+q)\log(n-d)/(n-d), \tag{1.8}$$

where $\hat{\sigma}^2_{p,q}$ is the maximum likelihood estimator of σ^2. The criterion estimates the orders (p, q) by selecting $(\hat{p}, \hat{q})$ which minimizes (1.8).

The method just described to select the orders, which is based on the traditional BIC criterion, is computationally expensive because one has to perform a nonlinear optimization for each (p, q) to compute $\hat{\sigma}^2_{p,q}$. For this reason, Hannan and Rissanen (1982) propose to perform the estimation using linear regression techniques in three steps, although the third step is used to compute estimators of the ARMA model selected by the BIC criterion which have similar properties to the maximum likelihood estimators. Therefore, only the first two steps are used to select the orders (p, q).

Computation of $\mathrm{BIC}_{p,q}$

In the first step of the HR method, which takes place only if there is a moving average part ($q > 0$), estimates $\hat{a}_t$ of the innovations a_t in (1.7) are obtained by fitting a long autoregressive model to the series. That is, given a big positive integer N, the $\hat{a}_t$ are computed using

$$\hat{a}_t = \sum_{j=0}^{N} \hat{\phi}_N(j) w_{t-j}, \quad \hat{\phi}_N(0) = 1, \quad t \geq 1,$$

where $w_t = 0$ if $t \leq 0$ and the $\hat{\phi}_N(j)$ are computed using Durbin-Levinson's algorithm. This last algorithm consists of first estimating the sample auto-

covariances

$$c_t = \frac{1}{n-d} \sum_{s=1}^{n-d-t} w_s w_{s+t}$$

and then recursively computing the $\hat{\phi}_N(j)$ using the equations

$$\hat{\phi}_N(N) = -\sum_{j=0}^{N-1} \hat{\phi}_{N-1}(j) c_{N-j} / \hat{\sigma}^2_{N-1}, \quad \hat{\phi}_N(j) = \hat{\phi}_{N-1}(j) + \hat{\phi}_N(N) \hat{\phi}_{N-1}(N-j),$$

$$\hat{\sigma}^2_N = \{1 - \hat{\phi}^2_N(N)\} \hat{\sigma}^2_{N-1}, \quad \hat{\phi}_1(1) = c_1/c_0, \quad \hat{\sigma}^2_0 = c_0.$$

In the procedure proposed by Gómez (1998), the value of N is selected to be $N = max\{[\log^2(n-d)], 2max\{p,q\}\}$, where (p,q) are the orders of the ARMA model for which the BIC is being computed and $[\log^2(n-d)]$ is the integer part of $\log^2(n-d)$. This choice is based on the fact that Hannan and Rissanen (1982), p. 88, assume that n is greater than $\log(n-d)$, but not greater than $\log^b(n-d)$, for some $b < \infty$.

In the second step of the HR method, given the orders (p,q), first the parameters of model (1.7) are estimated by minimizing

$$S(p,q) = \sum_{t=m}^{n-d} \Big\{ \sum_{j=0}^{p} \phi_j w_{t-j} - \sum_{j=1}^{q} \theta_j \hat{a}_{t-j} \Big\}^2, \tag{1.9}$$

where $m = max\{p+1, q+1\}$ and $\phi_0 = 1$. Then, the estimator $\hat{\sigma}^2_{p,q}$ is computed by the formula $\hat{\sigma}^2_{p,q} = S(p,q)/(n-d)$ and the $\mathrm{BIC}_{p,q}$ statistic is computed using (1.8). The use of an efficient numerical method, like, for example, the application of the QR algorithm based on Housholder transformations, to minimize (1.9) is important to avoid singularity problems when both p and q are overspecified.

In the procedure proposed by Gómez (1998), the following modifications are made. If there is no moving average part ($q = 0$), the estimation of the parameters of the ARMA model finishes here. Note that, in this case, the estimates obtained for the autoregressive part coincide with the ones obtained by OLS.

If there is a moving average part ($q > 0$), the estimators $\tilde{\phi}_j$ and $\tilde{\theta}_j$, obtained by minimizing (1.9), are consistent but have a bias and, therefore, they are not asymptotically efficient. In order to obtain bias-corrected, consistent and asymptotically efficient estimators, see Zhao-Guo (1985), first form

$$\tilde{a}_t = -\sum_{j=1}^{q} \tilde{\theta}_j \tilde{a}_{t-j} + \sum_{j=0}^{p} \tilde{\phi}_j w_{t-j}, \quad t \geq 1,$$

where $\tilde{a}_t = 0$ and $w_t = 0$ if $t \leq 0$. Then put

$$\eta_t = -\sum_{j=1}^{p} \tilde{\phi}_j \eta_{t-j} + \tilde{a}_t, \quad \xi_t = -\sum_{j=1}^{q} \tilde{\theta}_j \xi_{t-j} + \tilde{a}_t, \quad t \geq 1,$$

where $\eta_t = 0$ and $\xi_t = 0$ if $t \leq 0$. Finally, regress $\tilde{a}_t$ on $-\eta_{t-j}$, $j = 1, \ldots, p$, and $\xi_{t-j}, j = 1, \ldots, q$. The estimated regression coefficients are added to the estimators $\tilde{\phi}_j$ and $\tilde{\theta}_j$ to obtain the desired estimators $\hat{\phi}_j$ and $\hat{\theta}_j$.

When there are a moving average ($q > 0$) and an autoregressive ($p > 0$) part, Gómez (1998) proposes to obtain better estimates of the moving average part by repeating the previous procedure with the series filtered with the autoregressive filter. That is, the series is first filtered with the autoregresive filter $\hat{\phi}(B)$ estimated in the two previous steps to obtain the series $x_t = \hat{\phi}(B)w_t$. Then, the series x_t, which asymptotically follows the model $x_t = \theta(B)a_t$ and, therefore, does not have an autoregressive part, is subject to the two previous steps.

Once the parameter estimates of model (1.7) have been obtained for some orders (p, q), the estimator $\hat{\sigma}^2_{p,q}$ is needed to compute the $\text{BIC}_{p,q}$ statistic. In the procedure proposed by Gómez (1998), the residuals r_t, $t = 1, \ldots, n = max\{p, q\}$ of the series w_t are first computed using a fast Kalman filter routine based on the algorithm of Morf, Sidhu and Kailath (1974). Then, the rest of the residuals r_t, $t = n+1, \ldots, n-d$ are recursively obtained using the difference equation (1.7). Finally, the estimator $\hat{\sigma}^2_{p,q}$ is computed by the formula

$$\hat{\sigma}^2_{p,q} = \frac{1}{n-d} \sum_{t=1}^{n-d} r_t^2,$$

and the $\text{BIC}_{p,q}$ statistic is computed using (1.8).

Optimization of $\text{BIC}_{p,q}$

After having described the algorithm to compute $\text{BIC}_{p,q}$ for each (p, q), we now review the algorithms used by the HR method and he procedure proposed by Gómez (1998) to obtain the optimal model of the form (1.7). In the HR method, the model is selected as that $\text{ARMA}(\tilde{p}, \tilde{q})$ model for which $\text{BIC}_{\tilde{p},\tilde{q}}$ is minimum among all $\text{ARMA}(p, q)$ models satisfying $p \leq P$ and $q \leq Q$, where P and Q are fixed upper bounds. These authors recommend to search first among models with $p = q$ and refine the search later.

To describe the procedure proposed by Gómez (1998), suppose the general case, where the series follows a multiplicative seasonal model given by (1.4).

In practice, it is assumed that the orders of the ARMA$(p_r, q_r)\times(p_s, q_s)_s$ model followed by the series verify $0 \leq p_r, q_r \leq 3$ and $0 \leq p_s, q_s \leq 2$, and the BIC statistic should be computed for all these combinations. Since the resulting number of combinations is high, the search is performed sequentially. The algorithm is:

I) Specify first an ARMA$(3,0)$ model for the regular part. Then, compute the BIC statistic for models where the seasonal part verifies $0 \leq p_s, q_s \leq m_s$, and select the minimum. The number m_s is selected by the user, the default value being 1.

II) Fix the seasonal part to that selected in I), compute the BIC statistic for models where the regular part verifies $0 \leq p_r, q_r \leq m_r$, and select the minimum. The number m_r is selected by the user, the default value being 3.

III) Fix the regular part to that selected in II), compute the BIC statistic for models where the seasonal part verifies $0 \leq p_s, q_s \leq m_s$, and select the minimum. The number m_s is that of I).

The justification for the previous algorithm is as follows. In I), the regular part is assumed to be an ARMA$(3,0)$ model. This is usually a good approximation to many regular models found in practice, so that step I) amounts to first filtering the series with the approximate regular model and then finding a seasonal model for the filtered series. This seasonal model will probably be a good approximation to the seasonal part. In step II), we filter the series with the seasonal model found in I) and find an appropriate regular model. In step III), we filter the series with the regular model found in II) and look for an appropriate seasonal model. Clearly we could iterate this procedure further, but usually the three steps are enough to find a satisfactory model.

The previous algorithm allows for a substantial reduction in computing time and, however, the results obtained with it are very satisfactory. Once the previous algorithm has finished, and in order to avoid the tendency of BIC to overparametrize, especially in the seasonal part, the smallest five BIC are first ordered in ascending order. Then, the first one is compared to the other four and if the difference in absolute value is less than a certain number and the biggest of the two BIC corresponds to a more parsimonious seasonal part, this last one is selected. Among all the BIC that satisfy this condition, the one that corresponds to the more parsimonious part is selected, provided that the seasonal part exists ($p_s > 0$ or $q_s > 0$). The procedure also favours

balanced models (models where the degrees of the autoregressive and the moving average parts coincide).

In the previous algorithm, if the parameters estimated for an ARMA model are such that the roots of the autoregressive or the moving average polynomials lie within the unit circle, this fact is considered as an indication of model inadequacy and the model is rejected.

The tentative model ARMA$(3,0)$ specified in I) of the previous algorithm seems to be robust and the sequential search of the algorithm has given very satisfactory results in all performed tests of the proposed procedure, with real and simulated series.

If there is a mean or other regression effects in model (1.7), the procedure proposed by Gómez (1998) obtains first OLS estimators of the regression parameters. Then, these effects are subtracted from the differenced series before computing the parameter estimates of model (1.7) and also before computing the residuals r_t needed in the computation of $\hat{\sigma}^2_{p,q}$ and the BIC statistic.

1.3.4 Liu's Filtering Method

Recently, the SCA software package has incorporated a module for automatic ARIMA model identification, called "SCA-Expert". This module uses a procedure based on the filtering method proposed by Liu (1989) and certain heuristic rules. Briefly, this method consists of the following:

1. Examine first the sample autocorrelation functions (SACF) of z_t, $(1-B)z_t$, $(1-B^s)z_t$ and $(1-B)(1-B^s)z_t$ to assert the differencing orders and to see whether seasonality is present. After that, examine the SACF of the properly differenced series. If an obvious seasonal ARIMA model can be specified from the SACF, stop. Otherwise, go to the following step. Denote by y_t the differenced series.

2. If an obvious tentative model cannot be deduced from the SACF of y_t, estimate an intermediate model of the type ARMA$(1,1)\times(1,1)_s$. If no one of the autoregressive parameters has is close to 1, generate the series R_t and S_t, which are the result of filtering y_t with the ARMA$(1,1)_s$ and ARMA$(1,1)$ models that make up the intermediate model.

   If any of the autoregressive parameters is close to 1, then difference properly. After differencing, a new intermediate model of the same type is estimated and new R_t and S_t series are generated.

3. Use the SACF and sample partial autocorrelation functions, as well as the extended sample autocorrelation function, of R_t to identify an ARMA model adequate for the R_t series.

4. In order to identify a model for S_t, the SACF of S_t can be used. If a model is not clear for S_t, examine also the estimated parameters for the seasonal part in the intermediate model and use them to specify a model for S_t.

One problem with the previous algorithm is that the computerized specification of either the differencing orders or a seasonal or regular model from the SACF or the sample partial autocorrelation function does not seem at all clear. On the other hand, the idea of filtering the series with an approximate regular or seasonal model to find the other part of the model is a good one and usually gives satisfactory results in practice.

1.4 Automatic Modeling Methods in the Presence of Outliers

Many time series encountered in practice have outlying observations. These may be due to errors in the data, strikes, changes in regulations, etc. The presence of outliers can make extremely difficult the process of model identification. For this reason, any automatic model identification method has to incorporate some kind of outlier treatment.

In this section, we start first with the definition of an outlier. Here we proceed quickly because outliers have already been dealt with in chapter 6. Second, after examining some algorithms for outlier treatment, we review the method proposed by Gómez (1998) for automatic outlier detection and correction. It is pointed out that in the previous algorithms an exact filter should be used, instead of the inverse of the model, which is the filter usually applied in practice. Third, some estimation and filtering techniques are reviewed which are used to speed up the algorithms of the previous methods. Fourth, some reasons are given for the need to robustify automatic modeling methods. Finally, an algorithm is proposed for automatic model identification in the presence of outliers.

1.4.1 Outlier Definition

When analyzing time series data, it is not unusual to find outlying observations due to uncontrolled or unexpected interventions, like strikes, major changes in political or economic policy, the occurrence of a disaster, gross errors, and so forth. Since ARIMA models, which are frequently used in time series modeling, are designed to grasp the information of processes with a homogeneous memory pattern, the presence of outlying observations or structural changes may influence the efficiency and goodness of fit of these models. See, for example, Abraham and Box (1979), Chen and Tiao (1990), Tsay (1986), and Guttman and Tiao (1978).

The traditional approach to handle the problem of outliers, once it is assumed that a proper ARIMA model has been correctly identified for the series, consists of first identifying the location and the type of outliers and then use the intervention analysis proposed by Box and Tiao (1975). This procedure requires that a time series expert first examines the data and then, with the help of some time series software, analyses the sample autocorrelation and partial autocorrelation functions of the residuals, graphs of the series and the residuals, etc. For this reason, it is important to try to find some procedure which automates as much as possible the process of detection and correction of outliers. Among the first steps in this direction, we can mention the procedures of Chang, Tiao and Chen (1988), Hillmer, Bell and Tiao (1983), and Tsay (1988). These procedures are quite effective in detecting the locations and estimating the effects of large isolated outliers. However, the problem is not solved because

1) The presence of outliers may result in an incorrectly specified model

2) Even if the model is appropiately specified, outliers may still produce important biases in parameter estimates

3) Some outliers may not be identified due to a masking effect

The method proposed by Tsay (1986) is an important contribution to solve the problem of model identification in the presence of outliers. Chen and Liu (1993) have proposed a method for outlier treatment that tries to solve problems 2) and 3). This method works rather satisfactorily, although it presents some aspects that can be improved. These include: i) the method uses exact maximum likelihood estimation several times, thus making it computationally expensive; ii) it does not use exact residuals; iii) the algorithm is too complicated; iv) multiple regressions are performed without filtering

the data and the columns of the design matrix by an "exact" filter, like the Kalman filter. It uses instead a conditional filter. vi) The method uses a sort of backward elimination procedure to select the "best regression equation", instead of a stepwise procedure which is more robust.

The method proposed by Gómez (1998) for the detection and correction of outliers attempts to solve problems 2) and 3) in such a way that the above mentioned aspects in the procedure of Chen and Liu (1993) are improved, as will be described later. In addition, if it is used in the algorithm proposed at the end of this section, together with the algorithm for automatic model identification proposed by Gómez (1998), the proposed algorithm constitutes an alternative procedure to that of Tsay (1986) to solve the problem of automatic model identification in the presence of outliers that complements and may improve considerably Tsay's procedure.

Suppose first that there are no regression effects. We will extend the results to the general case later. Let the series $\{z_t\}$ follow the ARIMA(p,d,q) model given by (1.3), where it is understood that if the model is multiplicative seasonal, (1.4) holds. We will assume for simplicity that $\mu = 0$ and we will use the notation (1.3), referring to (1.4) when necessary. To model the effect of an outlier, consider the model

$$z_t^* = z_t + \omega\nu(B)I_t^T, \tag{1.10}$$

where $\nu(B)$ is a quotient of polynomials in B, $\{z_t\}$ is the outlier free series, $I_t^T = 1$ if $t = T$ and $I_t^T = 0$ otherwise, is an indicator function to refer to the time in which the outlier takes place and ω represents the magnitude of the outlier. Four types of outliers will be considered

IO: $\nu(B) = \theta(B)/(\delta(B)\phi(B))$,

AO: $\nu(B) = 1$,

TC: $\nu(B) = 1/(1-\delta B)$,

LS: $\nu(B) = 1/(1-B)$.

The acronyms stand for innovational outlier (IO), additive outlier (AO), temporary change (TC) and level shift (LS). The value of δ is considered fixed and is made equal to 0.7. These four types of outliers have already been considered in chapter 6. For additional information about the nature and motivation for these outliers, see Chen and Tiao (1990), Fox (1972), Hillmer, Bell and Tiao (1983), and Tsay (1988). These four outliers correspond to

some simple types of outliers. More complicated outliers can be usually approximated by combinations of these four types.

Innovational outliers may have a tremendous effect on the series level, especially for nonstationary series, due to the factor $\delta(B)$ in the denominator of the filter which defines the outlier. For this reason, the use of innovational outliers should be discouraged in routine application of automatic model identification procedures to many series. It is our practical experience that innovational outliers are often more a nuisance than a help when identifying and fitting ARIMA models.

1.4.2 Algorithms for Automatic Outlier Detection and Correction

We will start by considering that there is only one outlier. After having described how the effect of the outlier can be estimated and adjusted for, the case of multiple outliers will be considered. Finally, the algorithm proposed by Gómez (1998) will be reviewed. The emphasis here will be on exact filtering, as opposed to the usual practice of filtering with the inverse of the model followed by the series.

Estimation and Adjustment for the Effect of an Outlier

Suppose that the parameters in model (1.3) are known, the observed series is $z^* = (z_1^*, \ldots, z_n^*)'$, the outlier free series is $z = (z_1, \ldots, z_n)'$ and put $Y = (\nu(B)I_1^T, \ldots, \nu(B)I_n^T)'$. Then, (1.10) can be written as the regression model with ARIMA errors

$$z^* = Yw + z. \tag{1.11}$$

To simplify the exposition, we will assume that z in (1.11) follows an ARMA model or, what amounts to the same thing, $\delta(B) = 1$ in (1.3). If this is not the case, we would work with the series obtained by differencing z^*, z and Y in (1.11). Let $\mathrm{Var}(z) = \sigma^2\Omega$ and $\Omega = LL'$, with L lower triangular, the Cholesky decomposition of Ω. Premultiplying (1.11) by L^{-1}, the following ordinary least squares model is obtained

$$L^{-1}z^* = L^{-1}Yw + L^{-1}z. \tag{1.12}$$

Letting $r = L^{-1}z$, the equality $\mathrm{Var}(r) = \sigma^2 I_n$ holds and vector r is the residual vector of the series (not observed). If we let the estimated residuals be $r^* = L^{-1}z^*$ and write $X = L^{-1}Y$, (1.12) can be written as

$$r^* = Xw + r. \tag{1.13}$$

If Y is 0 in (1.11), the model would be $z^* = z$ and if we applied the Kalman filter to this model, we would obtain $L^{-1}z^*$. This result, which a standard result of control theory, allows us to see the Kalman filter as an algorithm that, applied to any vector v instead of z^*, yields $L^{-1}v$. Therefore, if we apply the Kalman filter to the vector of observations z^* and to the vector Y, we can move from (1.11) to (1.12) or, what amounts to the same thing, from (1.11) to (1.13).

We can estimate ω by OLS in (1.13) to obtain

$$\tilde{\omega} = (X'X)^{-1}X'r^*, \tag{1.14}$$

where the estimator variance is $\mathrm{Var}(\tilde{\omega}) = (X'X)^{-1}\sigma^2$. To test the null hypothesis that there is no outlier at $t = T$, we can use the statistic

$$\tau = (X'X)^{1/2}\tilde{\omega}/\sigma, \tag{1.15}$$

which is distributed $N(0,1)$ under the null.

In practice, the parameters of model (1.3) will not be known and they will have to be estimated. Under these circumstances, the usual procedure consists of estimating first the parameters of model (1.3) by exact maximum likelihood, as if there were no outliers, and then using instead of (1.14) and (1.15) their sample counterparts

$$\hat{\omega} = (\hat{X}'\hat{X})^{-1}X'\hat{r}^*, \qquad \hat{\tau} = (\hat{X}'\hat{X})^{1/2}\hat{\omega}/\tilde{\sigma},$$

which are obtained by replacing in (1.14) and (1.15) the unknown parameters with their estimates. It can be shown that $\hat{\tau}$ is asymptotically equivalent to τ. See Chang, Tiao and Chen (1988), p. 196. Each matrix X and, therefore, $X'X$ depends on the type of the outlier.

To see whether there is an outlier at $t = T$, the four estimators $\hat{\omega}_{IO}^T$, $\hat{\omega}_{AO}^T$, $\hat{\omega}_{TC}^T$ and $\hat{\omega}_{LS}^T$ are first computed, along with the statistics $\hat{\tau}_{IO}^T$, $\hat{\tau}_{AO}^T$, $\hat{\tau}_{TC}^T$ and $\hat{\tau}_{LS}^T$, where the subindex refers to the outlier type. Then, as proposed by Chang, Tiao and Chen (1988), the statistic $\lambda_T = max\{|\hat{\tau}_{IO}^T|, |\hat{\tau}_{AO}^T|, |\hat{\tau}_{TC}^T|, |\hat{\tau}_{LS}^T|\}$ is used. If $\lambda_T > C$, where C is a predetermined critical value, then there is the possibility of an outlier of the type given by the subindex of the statistic $\hat{\tau}$ for which the maximum is obtained.

Since the time $t = T$ at which the outlier occurs is unknown in practice, the criterion based on the likelihood quotient of Chang y Tiao (1983), leads to repeat the previous operation for $t = 1, \ldots, n$ and compute $\lambda = max_t\lambda_t = |\hat{\tau}_{tp}^T|$, where tp can be IO, AO, TC or LS. If $\lambda > C$, then there is an outlier of type tp at T.

Once the type of an outlier at $t = T$ is known, the series and the residuals can be corrected for its effect using (1.10) and (1.13).

Up to now, we have assumed that r^* and X were computed by means of an "exact" filter, which was the Kalman filter. This is the correct thing to do, since the number of observations in a time series is always finite and we cannot apply the semi-infinite filter, given by the inverse of the series model $\pi(B) = 1 + \pi_1 B + \pi_2 B^2 + \cdots = \phi(B)\delta(B)/\theta(B)$, to (1.10) to obtain

$$\pi(B)z_t^* = \omega[\pi(B)\nu(B)I_t^T] + a_t, \quad t = 1, \ldots, n,$$

instead of (1.12). In practice, the usual procedure consists of truncating the filter $\pi(B)$ and disregarding some observations at the beginning of the series. See Chen and Liu (1993), p. 285. In the procedure proposed by Gómez (1998), the residuals are filtered with an exact filter to obtain r^* and the filter $\pi(B)$ is used to filter the vector Y in (1.11). Note that using the Kalman filter to filter the vector Y for each possible combination of outliers would be computationally burdensome.

The Case of Multiple Outliers

When multiple outliers are present, we should use instead of (1.10) the model

$$z_t^* = z_t + \sum_{i=1}^{k} \omega_i \nu_i(B) I_t^{t_i}. \tag{1.16}$$

As shown by Chen and Liu (1993), the estimators of the ω_i obtained simultaneously using (1.16), can be very different from the ones obtained by an iterative process using the results of the previous section. That is, by obtaining first $\hat{\omega}_1$, then $\hat{\omega}_2$, etc. For this reason, it is important that every algorithm for outlier detection performs at some point multiple regressions to detect spurious outliers and correct the bias produced in the estimators sequentially obtained.

In order to estimate the parameters in the multiple regressions, when the parameters of the ARIMA model (1.3) are assumed to be known, the algorithm proposed by Gómez (1998) uses first the Kalman filter like when we moved from (1.11) to (1.12). Then, the estimators of the ω_i and the corresponding statistics are computed using (1.14) and (1.15). This is done in an efficient manner, using the QR algorithm and Housholder transformations. A more detailed description will be given at the end of this section.

Estimation of the Standard Deviation σ of the Residuals

When outliers are present in the series, the usual sample estimator can overestimate σ. For this reason, it is advisable to use a robust estimator. In the procedure proposed by Gómez (1998) the estimator used is the MAD estimator, defined by

$$\hat{\sigma} = 1.483 \times median\{|r_t^* - \tilde{r}^*|\},$$

where $\tilde{r}^*$ is the median of the estimated residuals $r^* = L^{-1}z^*$. The parameters of the model were assumed to be known in the previous formula. If they were unknown, they would be replaced with their estimates, as usual.

For outlier treatment, the procedure proposed by Gómez (1998) assumes that the orders (p, d, q) of model (1.3) are known and it proceeds iteratively. In the first stage, outliers are detected one by one and the model parameters are modified after each outlier has been detected. When no more outliers have been detected, the procedure goes to the second stage, where a multiple regression is performed. The outliers with the lowest t-value is discarded and the procedure goes back to the first stage to iterate.

The procedure used to incorporate or reject outliers is similar to the stepwise regression procedure for selecting the "best" regression equation. This results in a more robust procedure than that of Chen and Liu (1993), which uses "backward elimination" and may therefore detect too many outliers in the first stage of the procedure.

Up to now, we have supposed that there were no regression effects, but it is easy to incorporate these effects into the procedure. Let the series follow the regression model with ARIMA errors

$$z_t = y_t'\beta + \nu_t, \quad t = 1, \ldots, n, \tag{1.17}$$

where $\beta = (\beta_1, \ldots, \beta_k)'$ is the vector containing the regression parameters, which may include the mean as the first component, $\{z_t\}$ is the observed series, $\{y_t\}$ are the vectors containing the regression variables and $\{\nu_t\}$ follows the ARIMA model (1.3) with $\mu = 0$. Then, the algorithm proposed by Gómez (1998) for automatic detection and correction of outliers, described in detail, is the following:

Initialization

> If there are any regression variables in the model, included the mean, the regression coefficients are estimated by OLS and the series is corrected for their effects.

Stage I: Detection and estimation of outliers one by one

I.1) The ARIMA parameters are estimated, using the HR method, and the series is corrected for all regression effects present at the time, included the outliers so far detected. If desired by the user, exact maximum likelihood can be used for estimation, instead of the HR method.

I.2) Considering the estimates of the ARIMA parameters obtained in I.1 as fixed, the regression coefficients are estimated by GLS and their t statistics are computed. To this end, the fast algorithm of Morf, Sidhu and Kailath (1974) is used, followed by the QR algorithm. New estimated residuals are obtained.

I.3) With the estimated residuals obtained in I.2, the robust MAD estimator of the standard deviation of the residuals is computed.

I.4) If $u = (u_{d+1}, \ldots, u_n)'$, where d is the degree of the differencing operator, denotes the differenced series, the statistics $\hat{\tau}^t_{IO}$, $\hat{\tau}^t_{AO}$, $\hat{\tau}^t_{LS}$ and $\hat{\tau}^t_{TC}$ are computed for $t = d+1, \ldots, n$. To this end, the residuals computed in I.2 and the MAD obtained in I.3 are used. Let, for each $t = d+1, \ldots, n$, $\lambda_t = max\{|\hat{\tau}^t_{IO}|, |\hat{\tau}^t_{AO}|, |\hat{\tau}^t_{TC}|, |\hat{\tau}^t_{LS}|\}$. If $\lambda = max_t \lambda_t = |\hat{\tau}^T_{tp}| > C$, where C is a pre-selected critical value, then there is a possible outlier of type tp at T. The subindex tp can be IO, AO, TC or LS. If no outlier has been found the first time the algorithm passes through this point, then stop. The series is free from outlier effects. If no outlier has been found, but it is not the first time that the algorithm passes through this point, then go to II.1. If, on the contrary, an outlier has been found, then correct the series for all regression effects, using the estimates obtained in I.2 and the last outlier coefficient estimate obtained while computing λ, and go back to I.1 to iterate.

Stage II: Multiple Regression

Using the estimates of the multiple regression and their t statistics obtained the last time the algorithm passed through I.2, check whether there are any outliers with a t statistic $< C$, where C is the same critical value than in I.4. If there aren't any, stop. If, on the contrary, there are some, then remove the one with the lowest absolute t–value and go back to I.2 to iterate.

1.4.3 Estimation and Filtering Techniques to Speed up the Algorithms

To estimate the regression parameters in model (1.16), when the autoregressive and moving average parameters of the ARIMA model are assumed to be known, the procedure proposed by Gómez (1998) uses the following algorithm. Let the observed series $z = (z_1, \ldots, z_n)'$ follow the regression model with ARIMA errors

$$z = Y\beta + u, \tag{1.18}$$

where $\beta = (\beta_1, \ldots, \beta_k)'$ is the vector containing the regression parameters, which may include the mean as the first component, Y is an $n \times k$ matrix of full column rank and u follows the ARIMA model (1.3) with $\mu = 0$, which is supposed to be known. After differencing z, the columns of Y and u in (1.18), it is obtained that

$$y = X\beta + v, \tag{1.19}$$

where $y = (y_{d+1}, \ldots, y_n)'$, X is an $(n-d) \times k$ matrix, the components of $v = (v_{d+1}, \ldots, v_n)'$ follow the ARMA model $\phi(B)v_t = \theta(B)a_t$ and it is assumed that the degree of the differencing polynomial $\delta(B)$ is d.

If $\mathrm{Var}(v) = \sigma^2\Omega$ and $\Omega = LL'$, with L lower triangular, is the Cholesky decomposition of Ω, then, premultiplying (1.19) by L^{-1}, it is obtained that

$$L^{-1}y = L^{-1}X\beta + L^{-1}v, \tag{1.20}$$

which is an OLS model. As described in Section 1.4.2, the Kalman filter can be applied to y and the columns of the X matrix to move from (1.19) to (1.20). The Kalman filter algorithm used is the fast algorithm of Morf, Sidhu and Kailath (1974). β can now efficiently estimated in (1.20) by means of the QR algorithm. This last algorithm produces an orthogonal matrix Q such that $Q'L^{-1}X = (R', 0)'$, where R is a nonsingular upper triangular matrix. Partitioning $Q' = (Q_1, Q_2)'$ conforming to $(R', 0)'$, one can move from (1.20) to

$$\begin{aligned} Q_1'L^{-1}y &= R\beta + Q_1'L^{-1}v \\ Q_2'L^{-1}y &= \quad + Q_2'L^{-1}v, \end{aligned}$$

from which $\hat{\beta} = R^{-1}Q_1'L^{-1}y$ and $\hat{\sigma}^2 = y'(L^{-1})'Q_2Q_2'L^{-1}y/(n-d-k)$ are easily obtained. The Q matrix is obtained by means of Housholder transformations.

1.4.4 The Need to Robustify Automatic Modeling Methods

The presence of outliers in the series can affect tremendously all automatic model identification procedures, starting with the specification of unit roots and ending with the identification of an ARMA model for the differenced series. For this reason, there is a need to robustify automatic modeling methods. This can be achieved by the following scheme. Specify first a robust model for the series. This model could be the airline model, since, as mentioned earlier, it encompasses many models and is a model very often found in practice. Then, use this model to detect and correct the series for outlier effects. The critical value at this stage should not be low because we want to correct the series for the effects of the biggest outliers, which are the outliers that can distort most the automatic model identification procedure. With the series corrected for the outlier effects detected with the airline model, apply the automatic model identification procedure. With the model identified by this last procedure, specify a lower critical value and detect and correct the series for outliers. This cycle can be repeated several times until a satisfactory model is found. Usually, two iterations are enough.

1.4.5 An Algorithm for Automatic Model Identification in the Presence of Outliers

Taking into account the procedure proposed by Tsay (1986) and the previous considerations on how it could be improved, we propose an algorithmical procedure (implemented in programs TRAMO and SEATS, see Gómez and Maravall, 1997) which, briefly described, is the following:

a) *Preliminary tests.* If desired by the user, the procedure can test for the log-level specification, Trading Day and Easter effects. These last two tests are performed using the default model (airline model).

b) *Initialization.* If the user wants the series to be corrected for outliers, accept the model specified by the user (the default model is the airline model) and go to step 3. Otherwise, go to step 1. The critical value C for outlier detection can be either entered by the user or specified by the procedure. In this last case, the value of C is selected depending on the length of the series.

c) *Step* 1. If the user has specified the differencing orders and whether there should be a mean in the model, go to step 2. Otherwise, the

series is first corrected for all regression effects, if any. Then, using the corrected series, the differencing orders for the ARIMA model are automatically obtained and, also automatically, it is decided whether to specify a mean for the series or not. Go to step 2.

c) *Step* 2. Perform automatic identification of an ARMA(p, q) model for the differenced series, corrected for all outliers and other regression effects, if any. If the user wants to test for Trading Day and Easter effects and any of these effects was specified in the preliminary tests, check whether the specified effects are significant for the new model. If the user wants to correct the series for outliers, go to step 3. Otherwise, stop.

d) *Step* 3. Assuming the model known, perform automatic detection and correction of outliers using C as critical value. If a stop condition is not satisfied, perhaps decrease the critical value C and go to step 1.

In the previous algorithm, the procedures for obtaining the differencing orders, automatic model identification and automatic detection and correction of outliers are the ones proposed by Gómez (1998), which have been described in previous sections. The test for the log-level specification is the one considered in the previous section. The Trading Day and Easter effects, as well as tests for their presence in the model, will be described in detail in the next section.

1.5 An Automatic Procedure for the General Regression–Arima Model in the Presence of Outliers, Special Effects and, Possibly, Missing Observations

In this section, the algorithm for automatic model identification in the presence of outliers of last section is extended to the case in which there are missing observations. The algorithm was seen to handle any kind of regression effect. Special effects, like Trading Day and Easter effects are considered in detail, as well as intervention and other regression effects. Tests for the presence of Trading Day and Easter effects are given.

1.5.1 Missing Observations

The procedure proposed in the last section for automatic model identification in the presence of outliers can be extended easily to the case of missing observations. Missing observations are treated as additive outliers. This implies that we can work with a complete series, because the missing values are first assigned tentative values. Then, after the model has been estimated, the difference between the tentative value and the estimated regression coefficient is the interpolated value. See Gómez, Maravall and Peña (1998) for details.

Since we work with a complete series (there are no holes in it), we can use same algorithms described previously for automatic model identification and for automatic detection and correction of outliers. The tentative values assigned to the missing observations are the semisum of the two adjacent values.

1.5.2 Trading Day and Easter Effects

Traditionally, six variables have been used to model the trading day effect. These are: (no. of *Mondays*) - (no. of *Sundays*), ... (no. of *Saturdays*) - (no. of *Sundays*).

The motivation for using these variables is that it is desirable that the sum of the effects of each day of the week cancel out. Mathematically, this can be expressed by the requirement that the trading day coefficients β_j, $j = 1, \ldots, 7$, verify $\sum_{j=1}^{7} \beta_j = 0$, which implies $\beta_7 = -\sum_{j=1}^{6} \beta_j$.

Sometimes, a variable, called the length-of-month variable, is also included. This variable is defined as $m_t - \bar{m}$, where m_t is the length of the month (in days) and $\bar{m} = 30.4375$ is the average month length.

Another variable that can be used is the leap-year variable. This variable is equal to 0 for all months different from february. In february, it takes the value $-.25$ if february has 28 days, and .75 if february has 29 days (it is a leap year).

There is the possibility of considering a more parsimonious modeling of the trading day effect by using one variable instead of six. In this case, the days of the week are first divided into two categories: working days and non-working days. Then, the variable is defined as (no. of (M, T, W, Th, F)) - (no. of $(Sat, Sun) \times 5/2$).

Again, the motivation is that it is desirable that the trading day coefficients β_j, $j = 1, \ldots, 7$ verify $\sum_{j=1}^{7} \beta_j = 0$. Since $\beta_1 = \beta_2 = \ldots = \beta_5$ and $\beta_6 = \beta_7$, we have $5\beta_1 = -2\beta_6$.

The Easter variable models a constant change in the level of daily activity during the d days before Easter. The value of d is usually supplied by the user.

The variable has zeros for all months different from March and April. The value assigned to March is equal to $p_M - m_M$, where p_M is the proportion of the d days that fall on that month and m_M is the mean value of the proportions of the d days that fall on March over a long period of time. The value assigned to April is $p_A - m_A$, where p_A and m_A are defined analogously. Usually, a value of $m_M = m_A = 1/2$ is a good approximation.

Since $p_A - m_A = 1 - p_M - (1 - m_M) = -(p_M - m_M)$, the sum of the effects of both months, March and April, cancel out, a desirable feature.

Since Trading Day and Easter effects are modeled by means of regression variables, a possible test for these effects is the following. If no model has been identified, specify an airline model with mean. Otherwise, use the identified model. Then, using the differenced series y, apply first the Kalman filter to move from model (1.19) to model (1.20), where β is the vector of regression parameters, that includes the Trading Day and/or Easter parameters. Since model (1.20) is an OLS model, we can use an ordinary F-test to test if all Trading Day parameters are zero or not. A t-test can be used to test if the Easter parameter is zero.

1.5.3 Intervention and Regression Effects

Intervention variables are regression variables that are used to model certain abnormal effects, like strikes, major changes in economic policy, natural disasters, etc. See Box and Tiao (1975).

Examples of intervention variables are

- impulses
- level shifts
- temporary changes
- ramps

These variables usually consist of sequences of ones and zeros. Other regression effects, like economic variables thought to be related to the observed series, can also be incorporated.

1.6 Examples

The automatic model identification procedure proposed by Gómez (1998) and described earlier in this chapter was applied to 35 series which follow models covering a very broad spectrum. The TRAMO program, which, as mentioned earlier, implements the automatic model identification and automatic outlier detection procedures proposed by Gómez (1998), was applied with the parameters "$IDIF = 3$, $INIC = 3$" specified in the input file. This means that "the program will search first for regular differences up to order 2 and for seasonal differences up to order 1. Then, it will continue with the identification of an ARMA model for the differenced series, searching for regular polynomials up to order 3 and for seasonal polynomials up to order 1". The test for the log-level specification was not applied, so that the parameter "LAM" was set to 1 (no logs) whenever necessary. The default value of "LAM" is 0 (logs). Also, the parameter "MQ", which is the seasonal period, was set to the appropriate value whenever the seasonal period was different from 12, the default value.

The results are reported in Appendix A. Of the 35 series, 13 are series which have appeared in published articles and for which an ARIMA model has been identified by some expert in time series analysis. The rest are simulated series. For the simulated series, the identified models coincide with the models from which the series were generated. For the real series, TRAMO identifies either the same model as the one identified by the time series expert or an also acceptable, sometimes better, model.

In order to illustrate the use of the algorithm proposed by Gómez (1998) for automatic detection and correction of outliers, we consider the example of the ozone (O_3) mean levels in Los Angeles city during the period of January 1955 to December 1972. This series was analyzed by Box and Tiao (1975) as an example of a series for which intervention analysis could be applied.

Box and Tiao (1975) identified three intervention variables and a multiplicative moving average model for the series differenced with seasonal difference. More specifically, the model is

$$\begin{aligned} z(t) = \; & \omega_1 \; INT1(t) + \frac{\omega_2}{1 - B^{12}} INT2S(t) \\ & + \; \frac{\omega_3}{1 - B^{12}} INT2W(t) + \frac{(1 + \theta_1 B)(1 + \theta_{12} B^{12})}{1 - B^{12}} a(t), \end{aligned}$$

where $INT1$ is 1 in January 1960 and the following months and 0 otherwise, $INT2S$ is 1 in the summer months, starting in June 1966, and 0 otherwise, and $INT2W$ is 1 in the winter months, starting in 1966, and 0 otherwise.

Table 1.1: Outliers Identified for the Ozone Series

Outlier	Estimate	t–value	Type
$t = 11$	3.2773	4.82	AO
$t = 39$	-1.9287	-3.45	TC
$t = 21$	2.4878	3.71	AO
$t = 43$	-1.8824	-3.36	TC

The series, together with its intervention variables, was subject to the procedure proposed by Gómez (1998) for automatic detection and correction of outliers. The TRAMO program was applied with the parameters "*IATIP* = 1, *IMVX* = 1, *VA* = 3." specified in the input file. This specification means the following: i) the automatic outlier detection procedure will search for outliers of the 3 types, LS, AO and TC; ii) exact maximum likelihood will be used to estimate the parameters of ARIMA models during the outlier detection stage; iii) the critical level 3. will be used for the identification of outliers.

The results are displayed in Table 1.1. Four outliers have been identified. Two outliers of type AO, at $t = 11$ and $t = 21$, and two outliers of type TC, at $t = 39$ and $t = 43$.

To illustrate the algorithm of Section 1.4.5, we consider the example of the monthly variety stores sales considered by Hillmer, Bell and Tiao (1983). For the logged series, these authors identified the ARIMA model

$$\nabla\nabla_{12}z(t) = \frac{1 + \theta_{12}B^{12}}{1 + \phi_1 B + \phi_2 B^2}a(t). \tag{1.21}$$

The TRAMO program was run with the parameters "*LAM* $= -1$, *ITRAD* $= -1$, *IEAST* $= -1$, *IDIF* $= 3$, *INIC* $= 3$, *IATIP* $= 1$", specified in the input file. The first three parameters tell the program to perform the test for the log-level specification, trading day and Easter effect, respectively. The last parameter is used to specify automatic outlier detection. When this parameter is used in conjunction with *IDIF* $= 3$ and *INIC* $= 3$, the program will apply the algorithm of Section 1.4.5.

To implement the algorithm of Section 1.4.5, after performing the tests for the log-level specification, trading day and Easter effect, the TRAMO program can go through up to three rounds. In the first round, it uses the default model and default critical value C, or the model and critical value

Table 1.2: Outliers Identified for the Variety Stores Sales Series

Outlier	Estimate	t–value	Type
$t = 45$	.096	5.23	TC
$t = 96$	.084	-4.38	AO
$t = 112$	-.176	-10.18	LS

entered by the user, and detects and corrects the series for outliers. As mentioned earlier, the default model is the airline model of Box and Jenkins (1976). The default critical value C depends on the series length. In the second round, using the outlier corrected series, the program automatically identifies a model and, with that model, it performs a second automatic detection and correction of outliers. Usually, these two rounds are enough to identify a model with a good fit. If this is not the case, the program iterates. After the third round, it the fit is still not acceptable, the program specifies a general model. This general model is an ARMA$(3,1)(0,1)_s$ for the differenced series, where the differencing orders are the same of the last round. At some point of the procedure, the identified model is compared to the airline model and the model with the best fit is selected. This is done because the airline model is a robust model and departures from this model can be unstable.

Using the TRAMO program in the manner just described, the following results were obtained for the variety stores sales series. The test for the log–level specification specified the logarithmic transformation for the data. Neither trading day nor Easter effect was detected. In the first round, using the default model (the airline model) and a critical value $C = 3.5$, the program detected outliers at $t = 45$, of type TC, at $t = 96$, of type AO and at $t = 112$, of type LS. After correcting the series for the outlier effects, the program identified first the differencing polynomial $\delta(B) = \nabla\nabla_{12}$, without specifying a mean for the differenced process. Then, the program identified model (1.21). With this model, the program detected the same outliers than before, as can be seen in Table 1.2.

Appendix A: Summary of the Automatic Model Identification for 35 Real or Simulated Series

Series	Simulated models or manually identified models	Model obtained by TRAMO	Comments
(1) Maddala (1972) Grunfeld's inversion series (N=20)	$(1+\phi_1 B+\phi_2 B^2)z(t) = C+a(t)$	same as left	
(2) Hillmer, Bell and Tiao (1983) Clothing sales (N=153)	$\nabla\nabla_{12}z(t) = (1+\theta_1 B+\theta_2 B^2) \times(1+\theta_{12}B^{12})a(t)$	same as left	
(3) Hillmer, Bell and Tiao (1983) Hardware sales (N=155)	$\nabla\nabla_{12}z(t) = (1+\theta_1 B) \times(1+\theta_{12}B^{12})a(t)$	same as left	
(4) Hillmer, Bell and Tiao (1983) Variety stores sales (N=153)	$(1+\phi_1 B+\phi_2 B^2)\nabla\nabla_{12}z(t) = (1+\theta_1 B)a(t)$	same as left	
(5) Box and Tiao (1983) Ozone sales (N=216)	$\nabla_{12}z(t) = C+(1+\theta_1 B) \times(1+\theta_{12}B^{12})a(t)$	same as left	
(6) Box and Tiao (1983) CPI series (N=234)	$\nabla z(t) = C+ (1+\theta_1 B)a(t)$	$\nabla\nabla_{12}z(t) =(1+\theta_1 B) \times(1+\theta_{12}B^{12})a(t)$	(b)
(7) Chatfield and Prothero (1973) Monthly sales series (N=77)	$(1+\phi_1 B)\nabla\nabla_{12}z(t) = (1+\theta_{12}B)a(t)$	$\nabla_{12}z(t) = C +(1+\theta_1 B+\theta_2 B^2)a(t)$	(c)
(8) Hamilton and Watts (1978) Weekday coffee data (N=178)	$(1+\phi_1 B)\nabla z(t) = (1+\theta_5 B^5)a(t)$	$\nabla z(t) = C+ (1+\theta_1 B) \times(1+\theta_5 B^5)a(t)$	(a)
(9) Box and Jenkins (1976) Series A (N=197)	$\nabla z(t) = (1+\theta_1 B)a(t)$, or $(1+\phi_1 B)z(t) = C+(1+\theta_1 B^1)a(t)$	$\nabla z(t) = (1+\theta_1 B)a(t)$	(d)
(10) Box and Jenkins (1976) Series C (N=226)	$(1+\phi_1 B)\nabla z(t) = a(t)$, or $\nabla^2 z(t) = (1+\theta_1 B^1+\theta_2 B^2)a(t)$	$(1+\phi_1 B)\nabla z(t) = a(t)$	(d)
(11) Box and Jenkins (1976) Series E (N=100)	$(1+\phi_1 B+\phi_2 B^2)z(t) = C+a(t)$, or $(1+\phi_1 B+\phi_2 B^2+\phi_3 B^3)z(t) = C+(1+\theta_1 B^1)a(t)$	$(1+\phi_1 B+\phi_2 B^2)z(t) = C+(1+\theta_1)a(t)$	(f)

Series	Simulated models or manually identified models	Model obtained by TRAMO	Comments
(12) Box and Jenkins (1976) Series F (N=70)	$(1+\phi_1 B)z(t) = C + a(t)$,	same as left	
(13) Box and Jenkins (1976) Series G (N=144)	$\nabla\nabla_{12}z(t) = (1+\theta_1 B)$ $\times(1+\theta_{12}B^12)a(t)$,	same as left	
(14) Ljung and Box (1979) Simulated series (N=75)	$z(t) = (1+\theta_1 B)a(t)$	same as left	
(15) Tsay and Tiao (1984) Simulated series with AR complex unit roots (N=100)	$(1+\phi_1 B+\phi_2 B^2)\nabla^2 z(t) =$ $(1+\theta_1 B)a(t)$	same as left	
(16) Box and Tiao Simulated series R1 (N=150)	$z(t) = C+(1+\theta_1+\theta_2 B^2)a(t)$	same as left	
(17) Box and Tiao Simulated series R2 (N=162)	$(1+\phi_1+\phi_2 B^2)z(t) =$ $C + a(t)$	same as left	
(18) Box and Tiao Simulated series R3 (N=147)	$\nabla z(t) = C + (1+\theta_1 B)a(t)$	same as left	
(19) Box and Tiao Simulated series R4 (N=161)	$\nabla z(t) = (1+\theta_1 B+\theta_2 B^6)a(t)$	$(1+\phi_6 B^6)\nabla z(t) =$ $(1+\theta_1 B)a(t)$	(e)
(20) Box and Tiao Simulated series R5 (N=155)	$\nabla^2 z(t) =$ $(1+\theta_1 B+\theta_2 B^2)a(t)$	same as left	
(21) Box and Tiao Simulated series R6 (N=178)	$(1+\phi_1 B+\phi_2 B^2)\nabla z(t) =$ $a(t)$	same as left	
(22) Box and Tiao Simulated series R7 (N=149)	$(1+\phi_1 B)z(t) = C + a(t)$	same as left	
(23) Box and Tiao Simulated series R8 (N=148)	$(1+\phi_1 B)\nabla z(t) = C + a(t)$	same as left	

Series	Simulated models or manually identified models	Model obtained by TRAMO	Comments
(24) Box and Tiao Simulated series R9 (N=151)	$\nabla z(t) = (1 + \theta_1 B)a(t)$	same as left	
(25) Box and Tiao Simulated series R10 (N=146)	$\nabla z(t) = C + (1 + \theta_1 B + \theta_2 B^2)a(t)$	same as left	
(26) Box and Tiao Simulated series S1 (N=150)	$\nabla\nabla_{12} z(t) = (1 + \theta_1 B) \times (1 + \theta_2 B^{12})a(t)$	same as left	
(27) Box and Tiao Simulated series S2 (N=162)	$(1 + \phi_1 B)\nabla_{12} z(t) = (1 + \theta_1 B^{12})a(t)$	same as left	
(28) Box and Tiao Simulated series S3 (N=147)	$\nabla z(t) = (1 + \theta_1 B^{12})a(t)$	same as left	
(29) Box and Tiao Simulated series S4 (N=161)	$(1 + \phi_1 B^{12})z(t) = C + (1 + \theta_1 B)a(t)$	same as left	
(30) Box and Tiao Simulated series S5 (N=155)	$(1 + \phi_1 B)\nabla_{12} z(t) = C + a(t)$	same as left	
(31) Box and Tiao Simulated series S6 (N=178)	$\nabla\nabla_4 z(t) = (1 + \theta_1 B)a(t)$	same as left	
(32) Box and Tiao Simulated series S7 (N=149)	$\nabla^2(1 + \phi_1 B^6)z(t) = a(t)$	same as left	
(33) Box and Tiao Simulated series S8 (N=148)	$(1 + \phi_1 B + \phi_2 B^2)z(t) = C + (1 + \theta_1 B^6)a(t)$	same as left	
(34) Box and Tiao Simulated series S9 (N=151)	$\nabla_{12} z(t) = C + (1 + \theta_1 B)a(t)$	same as left	
(35) Box and Tiao Simulated series S10 (N=146)	$\nabla\nabla_{12} z(t) = (1 + \theta_1 B + \theta_2 B^{12})a(t)$	$\nabla\nabla_{12} z(t) = (1 + \theta_1 B) \times (1 + \theta_2 B^{12})a(t)$	(e)

(a) The model obtained by TRAMO is better, although the original model is also acceptable

(b) The model obtained by TRAMO is also acceptable. The seasonality is rather stable (θ_{12} = -.92223). Using the SEATS program, it can be verified that the

seasonality is also small and may be neglected.

(c) The model obtained by TRAMO is better. The model used in the original article is overdifferenced.

(d) In the original book, two alternative models were considered. TRAMO obtains the best one.

(e) TRAMO uses, in its automatic option, multiplicative models due to their simplicity and that they have less problems with nonstationarity and noninvertibility. However, by selecting a non-automatic option, the analyst may use non-multiplicative models if he prefers to do so.

(f) In the original book, two alternative models were considered. TRAMO obtains a model better than any of them.

References

1. Abraham, B. and Box, G. E. P., (1979), "Bayesian Analysis of Some Outlier Problems in Time Series", *Biometrika*, **66**, 229-236.

2. Beguin, J. M., Gourieroux, C. and Monfort, A, (1980), "Identification of a Mixed Autoregressive Moving Average Process: The Corner Method", in *Time Series*, O. D. Anderson, Ed., North–Holland, Amsterdam, 423–436.

3. Box, G.E.P., and Cox, D.R., (1964), "An Analysis of transformations", *Journal of the Royal Statistical Society*, Series B, **26**, 211–243.

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

5. Box, G.E.P., and Tiao, G.C., (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", *Journal of the American Statistical Association*, **70**, 70-79.

6. Brockwell, P., and Davis, R., (1992), *Time Series: Theory and Methods, Second Edition*, Springer-Verlag, Berlin.

7. Chang, I., Tiao, G.C. (1983), "Estimation of Time Series Parameters in the Presence of Outliers", Technical Report 8, University of Chicago, Statistics Research Center.

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

9. Chatfield, C., (1979), "Inverse Autocorrelations", *Journal of the Royal Statistical Society*, Series A, **142**, 363–377.

10. Chen, C. and Liu, L., (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association*, **88**, 284-297.

11. Chen, C. and Tiao, G. C., (1990), "Random Level Shift Time Series Models, ARIMA Approximation, and Level Shift Detection", *Journal of Business and Economic Statistics*, **8**, 170-186.

12. Choi, B., (1992), ARMA *Model Identification*, Springer Verlag, New York.

13. Cleveland, W.S., (1972), "The Inverse Autocorrelations of a Time Series and Their Applications", *Technometrics*, **14**, 277–298.

14. Dickey, D. A. and Pantula, S. G., (1987), "Determining the Order of Differencing in Autoregressive Processes", *Journal of Business and Economic Statistics*, **5**, 455–461.

15. Fox, A. J., (1972), "Outliers in Time Series", *Journal of the Royal Statistical Society*, Series B, **34**, 350-363.

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

17. Gómez, V., and Maravall, A., (1997), "Programs TRAMO and SEATS", Instructions for the User (Beta Version: June 1997), *Working Paper* 97001, Dirección General de Análisis y P.P., Ministerio de Economía y Hacienda, Madrid.

18. Gómez, V., Maravall, A., and Peña, D. (1998), "Missing Observations in ARIMA Models: Skipping Strategy Versus Additive Outlier Approach", *Journal of Econometrics* (Forthcoming).

19. Gray, H. L., Kelley, G. D. and McIntire, D. D., (1978), "A new Approach to ARMA Modeling", *Communications in Statistics*, **B7**, 1–77.

20. Guttman, I. and Tiao, G. C., (1978), "Effect of Correlation on the Estimation of a Mean in the Presence of Spurious Observations", *The Canadian Journal of Statistics*, **6**, 229-247.

21. Hamilton, D. C. and Watts, D. G., (1978), "Interpreting Partial Autocorrelation Function of Seasonal Time Series Models", *Biometrika*, **65**, 135-140.

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

23. Ljung, G. M. and Box, G. E. P., (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models", *Biometrika*, **66**, 265-270.

24. Liu, L. M., (1989), "Identification of Seasonal ARIMA Models Using a Filtering Method", *Communications in Statistics*, **18**, 2279-2288.

25. Liu, L. M. and Hanssens, D. M., (1982), "Identification of Multiple-Input Transfer Function Models", *Communications in Statistics*, **A11**, 297–314.

26. Lütkepohl, H., (1985), "Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process", *Journal of Time Series Analysis*, **6**, 35–52.

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

28. Morf, M., Sidhu, G.S., and Kailath, T., (1974), "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete- Time Systems", *IEEE Transactions on Automatic Control*, **AC-19**, 315-323.

29. Pearlman, J.G., (1980), "An Algorithm for the Exact Likelihood of a High-Order-Autore- gressive-Moving Average Process", *Biometrika*, **67**, 232-233.

30. Reinsel, G. C., (1997), *Elements of Multivariate Time Series Analysis*, second edition, Springer Verlag, New York.

31. Schwert, G. W., (1989), "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, **7**, 147–159.

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

33. Tsay, R. S., (1984), "Regression Models With Times Series Errors", *Journal of the American Statistical Association*, **79**, 118-124.

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

35. Tsay, R. S., (1988), "Outliers, Level Shifts, and Variance Changes in Time Series", *Journal of Forecasting*, **7**, 1-20.

36. Tsay, R. S. and Tiao, G. C., (1984), "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Functions for Stationary and Nonstationary ARMA models", *Journal of the American Statistical Association*, **79**, 84-96.

37. Tsay, R. S. and Tiao, G. C., (1985), "Use of Canonical Analysis in Time Series Model Identification", *Biometrika*, **72**, 299-316.

38. Zhao-Guo, C., (1985), "The Asymptotic Efficiency of a Linear Procedure of Estimation for ARMA Models", *Journal of Time Series Analysis*, **6**, 53-62.