20.II.2001

Seasonal outliers in time series

Regina Kaiser* and Agustín Maravall

Abstract

The standard procedures for automatic outlier detection and correction consider four types of outliers, namely, the additive, innovational, level shift, and transitory change outliers. In this paper, it is argued that this typification presents serious shortcomings. First, the innovational outlier may display undesirable features; second, it is incomplete because it cannot model breaks in the pattern of the seasonal component. Several specifications for a seasonal outlier are considered and the one denoted Seasonal Level Shift (SLS) is analysed in detail through simulation and real examples. It is concluded that the SLS displays better properties and turns out to be more useful than the innovational outlier, and hence the typification of outliers in automatic outlier detection and correction should replace the latter type of outlier by the seasonal level shift one.

Keywords: ARIMA Models, Seasonality, Level Shift, Outlier Detection and Correction.

Regina Kaiser
D. de Estadística y Econometría
Universidad Carlos III de Madrid
Madrid,126
28903 Getafe (Madrid)
Email: kaiser@est-econ.uc3m.es

Agustín Maravall
Sub. de Información y Métodos de Cálculo
Servicio de Estudios - Banco de España
Alcalá 50
28014 Madrid
Email: maravall@bde.es

* Partially supported by the Spanish grant PB95-0299 of CICYT

1 Introduction

Outlier detection has become an important part of time series analysis and influences modelling, testing and inference, because outliers can lead to model misspecification, biased parameter estimation, poor forecasts and inappropriate decomposition of the series (for example, poor seasonal adjustment). Outlier detection has become a key feature in automatic time-series model identification; examples are the REG-ARIMA routine in the X-12 ARIMA program (Findley et al., 1998); program TRAMO ("Time series Regression with Arima noise, Missing observations, and Outliers"; Gómez and Maravall, 1996), and the time series module in the Scientific Computing Associates package (Liu et al., 1986).

In automatic outlier detection and correction procedures, four outlier types are traditionally considered (see, for instance, Fox, 1972, Tsay, 1986, Chen and Liu, 1993 or Peña, 2001): *Additive Outliers* (AO), *Innovative Outliers* (IO), *Level Shift* (LS) outliers and *Transitory Change* (TC) outliers. These four outlier types affect an observed time series in different ways. The effect of an AO, an LS or a TC is independent of the ARIMA model for the series, while the effect of an IO on an observed series consists of an initial shock that propagates in the subsequent observations with the weights of the moving average representation of the ARIMA model. In so far as these weights are often explosive, the influence of an IO may, in some cases, increase as the date of its occurrence becomes more and more distant into the past, a rather undesirable feature. IOs are thus of a questionable interest. A further limitation of the standard approach is that the outlier typification is inadequate to capture possible breaks in the pattern of a seasonal component.

In this paper we suggest replacing IOs with some seasonal outlier in the typification of outliers used in automatic outlier detection and correction. Section 2 of the paper discusses the IO and some basic specifications for a seasonal outlier. Section 3 extends the automatic procedure to incorporate a seasonal outlier. Centering attention on seasonal level shifts, Section 4 uses simulation to analyse the performance of the extended procedure, and Section 5 looks at the effects of misspecifying a seasonal outlier. Finally, Section 6 discusses four real examples.

2 Outliers in ARIMA time series: some basic concepts

Let y_t be a time series that is the output of the ARIMA model

$$\phi(B)\delta(B)y_t = \theta(B)a_t, \tag{2.1}$$

where a_t is a niid (i.e., white noise) process with zero mean and variance V_a. B denotes the lag operator, such that $B^j y_t = y_{t-j}$, $\phi(B)$ and $\theta(B)$ are polynomials in B with all roots lying outside the unit circle, and $\delta(B)$ is the polynomial with unit roots that renders y_t stationary. For seasonal series, often the polynomials $\phi(B), \theta(B)$, and $\delta(B)$ display a multiplicative structure (Box and Jenkins, 1970, Ch. 9), so that, for example, $\delta(B) = \nabla^d \nabla_s^D$, where ∇ and ∇_s are the regular and seasonal differences, $\nabla = 1 - B$, $\nabla_s = 1 - B^s$, with s being the number of observations per year; stationarity, thus, is basically achieved by differencing. In practice, given an observed series y_t^*, prior to the ARIMA modelling stage, some modifications to the series are likely to be needed. Outlier correction is an essential one because the presence of one or more outliers in the observed series may seriously damage identification and estimation of the ARIMA model (see, for example, Chang, 1982).

Outliers denote observations that, broadly speaking, cannot be properly explained by the ARIMA model and its underlying normality assumption. They tend to be associated with irregular special events that produce a distortion on the series. In our approach, an important distinction is made between outliers and intervention variables (Box and Tiao, 1975). When there is a priori information about a special event that may have caused abnormal observations (the date of its occurrence and perhaps some idea of its likely effect, such as, for example, whether it is permanent or transitory,) the effect of the special event should be captured through intervention analysis. An outlier, on the other hand, represents anomalies in the observations for which no a priori information on the date of its occurrence or on the dynamic pattern of its effect are employed: they are revealed by analysis of the data. The main purpose of outlier correction is to modify the data in such a way that the normality hypothesis can be accepted, so that one can proceed with proper estimation, testing, and inference. Automatic outlier correction requires a prior decision on the types of dynamic patterns that will be considered, and a systematic procedure to detect them in the series. Because of their ad-hoc nature, it is desirable to keep the number of outliers as small as possible, unless they can be given convincing ex-post explanations.

Following the seminal work of Fox (1972), four different types of outliers have been proposed, together with several procedures to detect them (see, for instance, Tsay, 1986, Chen and Liu, 1993, Gómez and Maravall, 2001, and Kaiser, 1999). The outliers are classified as *Additive Outliers* (AO), *Innovative Outliers* (IO), *Level Shift* (LS) outliers or *Transitory Change* (TC) outliers. An AO represents an isolated spike, an LS a step function, a TC a spike that takes a few periods to disappear and an IO a shock in the innovations of the model. The four types of outliers are discussed in Peña (2001), and examples are displayed in Figure 1. Assuming, in general, that the observed series contains k outliers, their combined effect can be expressed as

$$y_t^* = \sum_{j=1}^{k} \xi_j(B)\omega_j I_t^{(\tau_j)} + y_t, \quad (2.2)$$

where y_t^* denotes the observed "contaminated" series; y_t follows the ARIMA process (2.1); ω_j is the initial impact of the outlier at time $t = \tau_j$; $I_t^{(\tau_j)}$ is an indicator variable such that it is 1 for $t = \tau_j$, and 0 otherwise; and $\xi_j(B)$ determines the dynamics of the outlier occurring at $t = \tau_j$, according to the following scheme:

$$\text{AO:} \quad \xi_j(B) = 1, \quad (2.3a)$$
$$\text{LS:} \quad \xi_j(B) = 1/(1-B), \quad (2.3b)$$
$$\text{TC:} \quad \xi_j(B) = 1/(1-\delta B), \quad 0 < \delta < 1, \quad (2.3c)$$
$$\text{IO:} \quad \xi_j(B) = \tilde{\theta}(B)/\tilde{\phi}(B). \quad (2.3d)$$

(For the TC outlier, the value δ is often preset at .7).

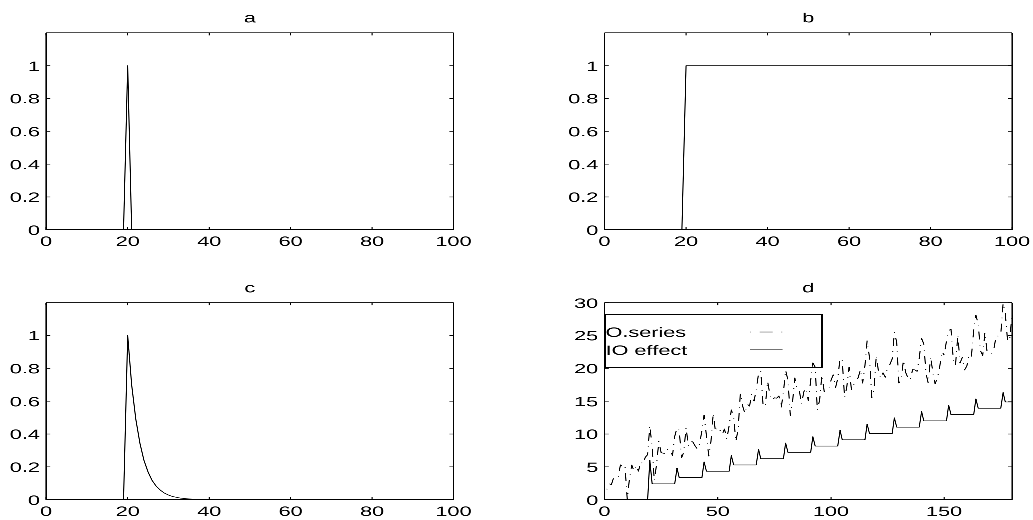

Figure 1. Effect of different types of outliers on the observed series a) AO, b) LS, c) TC and d) IO for an "Airline model" (see section 4).

The effect of an AO, TC, or LS outlier is independent of the ARIMA model for the series; for the AO and TC outlier, the effect is transitory, while for the LS outlier it is permanent. In all three cases the effect is bounded. On the other hand, the effect of an IO outlier depends on the particular model for the series, and for series with stationary transformation given by, for example, $\delta(B) = \nabla\nabla_{12}$, or $\delta(B) = \nabla^2$, the effect will be unbounded, as shown in Figure

1d. Figure 2 displays, for the example of Figure 1d, the percentage of the level of the series explained by the outlier, and it is seen that, after the initial impact, it increases as the date of the outlier becomes more and more distant into the past. Thus, in so far as the IO may have occurred in the relatively distant past, we would be living in a world that would be strongly determined by very old IOs; this is an undesirable property. It seems sensible to require that outliers should be specified in such a way that their effect remains, in some sense, bounded, and may not become a dominant explanation for the behavior of the series in some distant future.

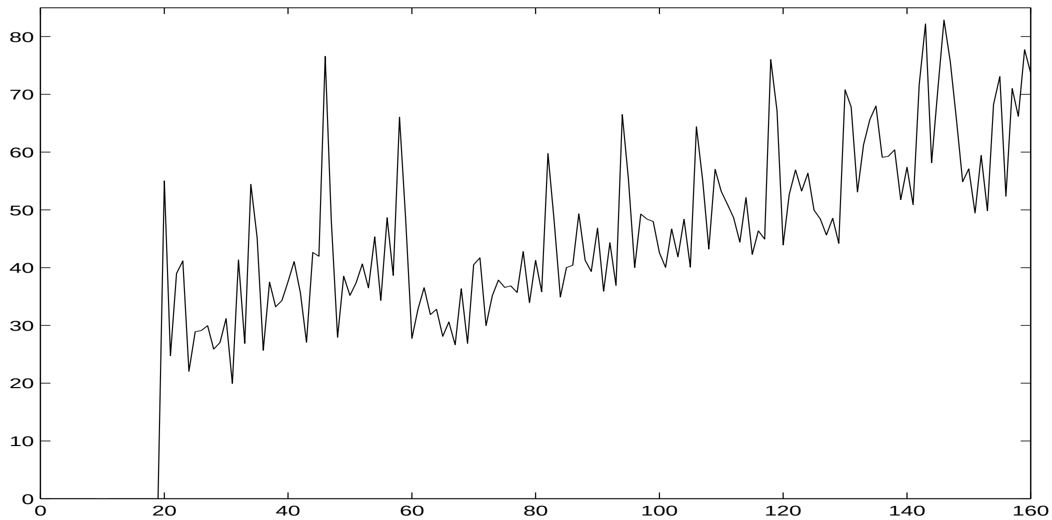

Figure 2. Percentage of the level of the series explained by the IO.

Further, for series that contain trend and seasonality, an IO will affect both, and Figure 3 exhibits the trend and seasonal decomposition of the IO of Figure 1d. But one can think of special events that produce breaks in the seasonal component of a series without producing at the same time an explosive break in the trend. The forced association of both effects is unnecessarily restrictive, and represents another undesirable feature of the IO (see also Peña, 1990). In summary, IOs present some serious drawbacks and, in our view, should be avoided. The problem then, is that the remaining three types of outliers cannot explain any changes in the seasonal component. This is an important limitation because, in our experience, seasonal breaks are often present in actual economic time series.

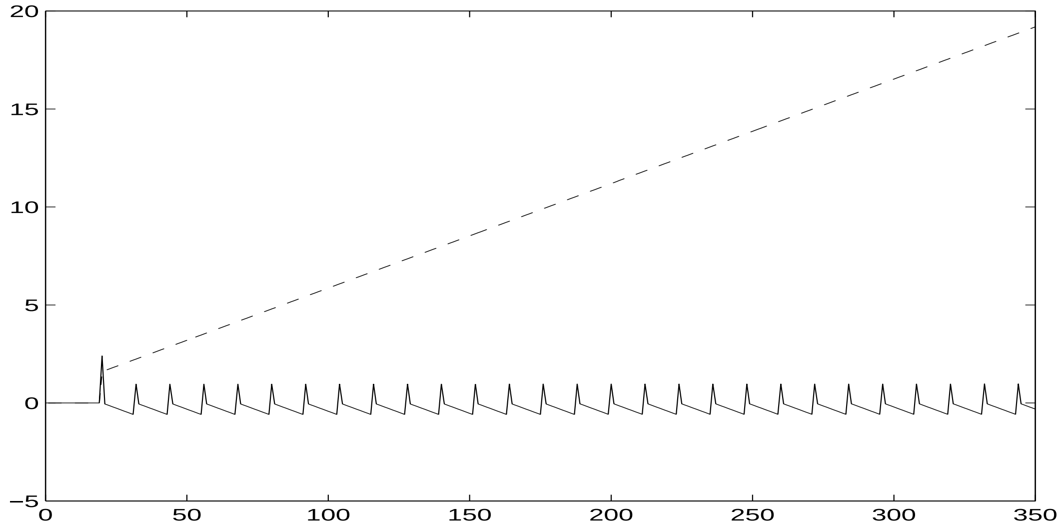

Figure 3. Trend and seasonal effects of an IO.

2.1 Seasonal outliers

Figure 4 displays the monthly Italian Industrial Production Index for the period 1981-1996. Using programs TRAMO and SEATS in the automatic mode (with RSA=4; see Gómez and Maravall, 1996), the ARIMA model identified for the series is of the type

$$\nabla\nabla_{12} y_t = (1 + \theta_1 B)(1 + \theta_{12} B^{12}) a_t, \tag{2.4}$$

and the estimate of the seasonal component is given in Figure 5. Direct inspection of this figure suggest that the August factor (the low peak in the figure) experiences an increase in 1988 (from a mean level of 46 to a mean level of 47.7) and again in 1994 (the mean level increasing to 49.3). We would like to be able to capture these increases in the mean level of August without having to necessarily impose the presence of an explosive deterministic trend, as would be implied by an IO in model (2.4).

Several basic structures for a seasonal outlier (SO) seem possible. If S denotes the annual aggregation operator, in terms of the representation (2.2), the simplest specification would be to set

$$\xi(B) = 1/S; \quad S = 1 + B + \ldots + B^{s-1}; \tag{2.5}$$

which would generate a purely seasonal outlier of the type displayed in Figure 6a. Each year, the effect of the outlier spreads over the same two months, the effect on the second month exactly cancelling the effect on the first one. An example could be the effect on a salary series of switching the payment of a special bonus from December to January.

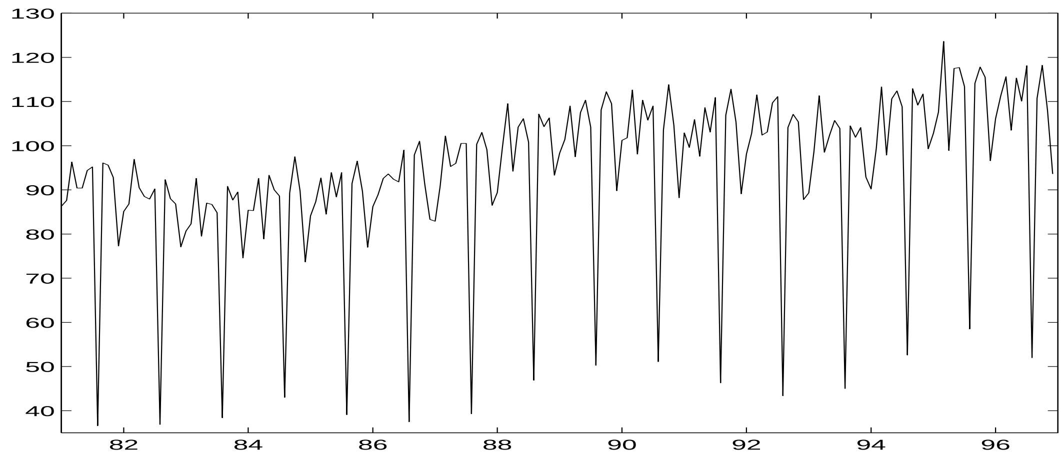

Figure 4. Italian industrial production index (GIPI).

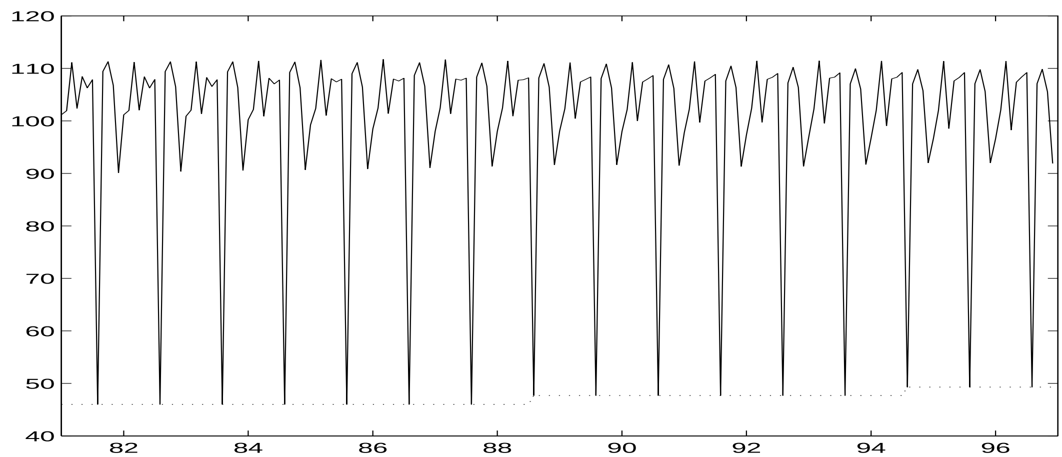

Figure 5. Italian industrial production index (GIPI). Seasonal factors

To fully cancel the effect in just one month may be inappropriate, and one may be interested in simply reflecting an increase (or decrease) for a given month. (An example could be a salary increase that will be paid as a special bonus in a particular month of each year.) One could then set

$$\xi(B) = 1/\nabla_s, \tag{2.6}$$

which generates the outlier of Figure 6b; we shall denote this outlier a Seasonal Level Shift (SLS). In this case, peaks for only one month are generated but, given that the outlier effect has a non-zero mean, it will also have an effect on the trend. Specifically, From the factorization $\nabla_s = \nabla S$, the polynomial S will affect the seasonal component, but the ∇ factor will affect the trend through a step increase in the mean. Therefore, the Seasonal Level Shift (SLS) outlier

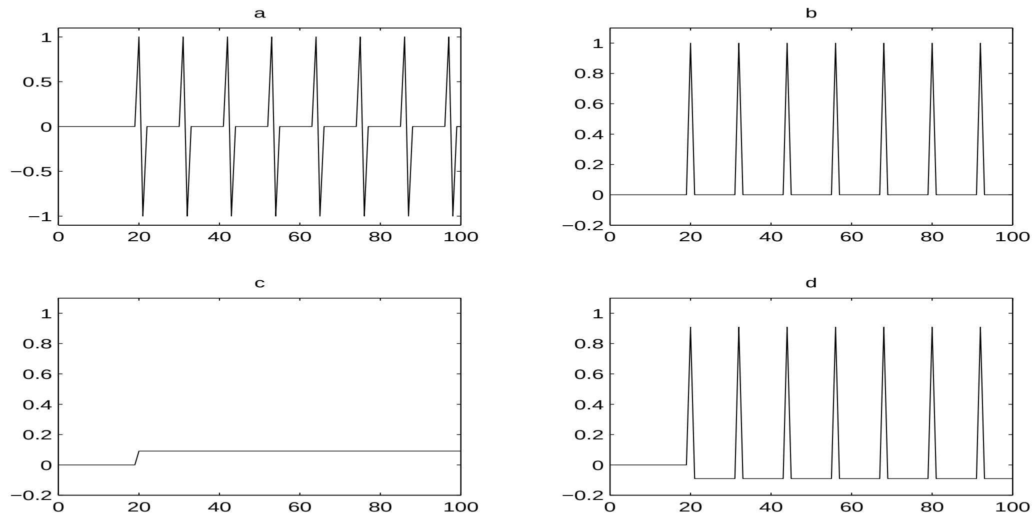

Figure 6. Seasonal Outliers effect. a) purely seasonal outlier b) Seasonal Level Shift c) effect of a SLS on the trend-cycle component d)effect of a SLS on the seasonal component

given by (2.6) has an effect on the trend given by the step function of Figure 6c, and an effect on the seasonal component shown in Figure 6d; both effects are permanent (due to the unit roots of ∇_s) though not explosive. Still, it may be that the effect on the trend of the SLS is not judged appropriate, and only the purely seasonal effect of Figure 6d is desired (see Bell 1983). This would reflect an effect in one month that is cancelled over the rest of the year, and an example could be the disappearance of a Christmas bonus that will now be paid uniformly over the year. In this case one can set

$$\xi(B) = \frac{1}{\nabla_{12}} - \frac{1}{12\nabla}, \tag{2.7}$$

or, equivalently, if $t = \tau$ denotes the starting period for the outlier effect, use the regression variable $z_t = 0$ for $t < \tau$, $z_t = 1$ for $t = \tau + sj$, and $z_t = -1/(s-1)$ for $t = \tau + sj + 1, \ldots, \tau + sj + s - 1$, for $j = 0, 1, 2, \ldots$

Naturally, more complex specifications can be employed, such as, for example, assuming that the spike affects more than one month in a year, or using specifications of the type $(1 - \alpha B)/\nabla_s$, or $(1 - \alpha B^s)/\nabla_s$ (see, for example, Kaiser and Maravall, 2000). However, given that our interest centers on a simple specification that could be used to complete the set of AO, TC and LS outliers in an automatic outlier detection procedure, we shall consider the three basic specifications (2.5), (2.6), and (2.7). All three outliers imply that variability of the seasonal component for some months is higher than for other months. There can be ways to deal with these different seasonal variabilities that are not based on the specification of outliers. For example, in an"ad-hoc" context, the program X12ARIMA permits the user to specify different seasonal filters for different months. In the ARIMA-model-based approach, this solution would be similar to using different seasonal moving average parameters for different months. But doing so, the model-based structure would either disappear or, at best, become considerably more complicated. From a structural time series model-based perspective (Harvey, 1989), Proieti (1998) models seasonal variability through heteroscedastic innovations in the seasonal component. Although ingenious, the approach is far from ready for incorporation into an automatic detection procedure.

Although the three basic specifications reflect different effects, it is of interest to see whether, for standard series length, fitting features could reliably distinguish between them. To look at this issue, we simulated 5000 series with model (2.4), with $\theta_1 = \theta_{12} = -.6$, and length=144 observations (a not too small, nor too large series, with a relatively stable seasonal component). Each series was perturbed at t=73 with, first, an outlier of the type (2.5), second, an outlier of the type (2.6), and, finally, an outlier of the type (2.7). In all cases, the parameter ω in (2.2) was set at $\omega = 3$ which implied that the t-value for its estimator was approximately t=4, and hence detectable with the usual critical values used in automatic outlier detection (see, for example, Chen and Liu, 1993). Then, for each series, model (2.4) was estimated under the assumption that there was an outlier of the type (2.5), (2.6), and (2.7), successively. Tables 1, 2 and 3 present the mean values and the standard deviations obtained for the Bayesian Information Criterion (BIC), the residual standard error ($SE(\hat{a}_t)$), and the Ljung-Box Q-statistics for the 36 residual autocorrelations. The elements in the main diagonal of each table represent the cases in which estimation is made with the correct specification. While the specification (2.5) implies an effect that only affects 2 months, the other two specifications affect all months in the year. Between these two, the only difference is the step increase in the mean, equal to $\omega/12$ for monthly data and to $\omega/4$ for quarterly one, and hence likely to be of a moderate size. Therefore, one could expect that the sample information would distinguish between the specification (2.5), on the one hand, and specifications (2.6) or (2.7), on the other, more easily than it would distinguish between the latter two specifications. Tables 1 to 3 confirm this expectation. Table 1 displays the residual root mean square error (RMSE ($\hat{a}_t$)) of the fit, Table 2 presents the Box-Ljung Q-statistics for 36 residual autocorrelations, and Table 3 the Bayesian Information Criterion (BIC), for the 16 possible combinations (outlier present-outlier fit).

First, it is seen that, for the 3 tables, the diagonal elements for the 3 SOs are practically identical (something to be expected, given that they correspond to the cases in which the correct outlier is fit). Second, it is seen that the effect of the outlier misspecification on the Q-statistics is clearly negligeable (besides, the values obtained are extremely close to the associated asymptotic χ^2 value with 22 degrees of freedom). Third, when there is no outlier, yet a SO is fit, the damage is also negligeable. But, when the series contains a SO but the outlier is ignored, there is a relatively small increase in RMSE ($\hat{a}_t$), and a large deterioration in the BIC statistics. Fourth, if a SO outlier of the type (2.6) is misspecified as in (2.7), the results are hardly affected. The same thing happens if an outlier of the type (2.7) is misspecified as in (2.6). However, if the outlier is specified as in (2.5), the BIC statistics deteriorates strongly.

In summary, specifications (2.6) and (2.7) for the seasonal outlier appear to produce very similar results. Both specifications imply that there is a big effect on one month, and a small uniform effect for the rest of the year. The specification (2.5), which implies back-to-back cancellation of the effect, cannot be taken as a good approximation to the former.

	Outlier in model				Aprox.
O.Estimated	No	(2.6)	(2.5)	(2.7)	SD
No	.99	1.11	1.22	1.10	.06
(2.6)	.99	.99	1.11	.99	.06
(2.5)	.99	1.05	.99	1.04	.06
(2.7)	.99	.99	1.10	.99	.06

Table 1. Mean value and standard deviation for residual standard error

	Outlier in model				Aprox.
O.Estimated	No	(2.6)	(2.5)	(2.7)	SD
No	21.8	21.2	21.9	21.2	7
(2.6)	21.9	21.9	21.5	21.9	7.2
(2.5)	21.9	21.8	21.9	21.8	7.2
(2.7)	21.9	21.9	21.5	21.9	7.2

Table 2. Mean value and standard deviations for Q-Statistic

	Outlier in model				Aprox.
O.Estimated	No	(2.6)	(2.5)	(2.7)	SD
No	7.7	38.4	65.6	37.8	17
(2.6)	12.2	12.2	43.4	12.5	17
(2.5)	12.2	27.0	12.2	26.5	17
(2.7)	12.2	12.4	42.7	12.2	17

Table 3. Mean value and standard deviations for BIC

3 Detection and estimation

In this section we detail how the iterative detection procedure described in Tsay (1986) and Gómez and Maravall(1994), to be denoted ”standard procedure”, can be enforced and extended to allow for both detection and estimation of a seasonal outlier (”extended procedure”).

3.1 A single outlier

Let α be the vector of parameters in model (2.1) and let us suppose, for the moment, that it is known. Further, suppose that the observed series is subject to the influence of a perturbation at time $t = \tau$ such that,

$$y_t^* = \xi(B)\omega I_t^{(\tau)} + y_t, \tag{3.1}$$

where we first assume that the model $\phi(B)y_t = \theta(B)a_t$ is stationary. Model (3.1) can be rewritten as a linear regression model as follows,

$$y_t^* = Z_t^*(\tau)\omega + y_t \tag{3.2}$$

where $Z_t^*(\tau) = \xi(B)I_t^{(\tau)}$ is an $N \times 1$ vector. Let $\mathbf{y}^* = (y_1^*, \ldots, y_N^*)'$; $\mathbf{y} = (y_1, \ldots, y_N)'$ and $\mathbf{Z}^* = (Z_1^*(\tau), \ldots, Z_N^*(\tau))'$. Writing (3.2) in matrix terms yields,

$$\mathbf{y}^* = \mathbf{Z}^*\omega + \mathbf{y}. \tag{3.3}$$

Assuming, first, that y_t is stationary, the model in (3.3) is a regression model with autocorrelated residuals and, therefore, the problem of estimating ω can be solved by Generalized Least Squares (GLS). Let $var(\mathbf{y}) = V_a\mathbf{\Omega}$ with $\mathbf{\Omega}$ a positive definite $N \times N$ matrix that depends on α and let $\mathbf{\Omega} = \mathbf{L}'\mathbf{L}$ be the Cholesky decomposition of $\mathbf{\Omega}$ with $\mathbf{L}$ lower triangular. Premultiplying (3.3) by $\mathbf{L}^{-1}$, and setting $\mathbf{e}^* = \mathbf{L}^{-1}\mathbf{y}^*$, $\mathbf{Z} = \mathbf{L}^{-1}\mathbf{Z}^*$ and $\mathbf{e} = \mathbf{L}^{-1}\mathbf{y}$, we obtain the Ordinary Least Squares (OLS) model,

$$\mathbf{e}^* = \mathbf{Z}\omega + \mathbf{e}, \tag{3.4}$$

where $var(\mathbf{e}) = V_a\mathbf{I}_N$. The OLS estimator of ω and its variance are obtained from (3.4) as,

$$\hat{\omega} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{e}^* \qquad var(\hat{\omega}) = (\mathbf{Z}'\mathbf{Z})^{-1}V_a. \tag{3.5}$$

As argued in Gómez and Maravall (1994), to move from the GLS model in (3.3) to the OLS model in (3.4), there is no need to evaluate the matrix $\mathbf{\Omega}$, since the application of the Kalman filter on the observed series $\mathbf{y}^*$. yields the vector of standardized residuals $\mathbf{e}^* = \mathbf{L}^{-1}\mathbf{y}^*$. Similarly, the application of the same filter on vector $\mathbf{Z}^*$ provides the vector $\mathbf{Z} = \mathbf{L}^{-1}\mathbf{Z}^*$ from which (3.5) can be computed.

To test the null hypothesis that the observation at time $t = \tau$ is not an outlier, one can use the standardized statistic,

$$\lambda = \frac{\hat{\omega}}{\sqrt{var(\hat{\omega})}} \tag{3.6}$$

which, for known α, follows asymptotically a standard normal distribution. By setting appropriate starting conditions (see, for example, Khon and Ansley, 1985, Bell and Hillmer, 1991, or Gómez and Maravall, 1994), the previous scheme extends in a straightforward manner to nonstationary series.

If the objective of the analysis is to determine the type of the outlier at time $t = \tau$, one possibility, suggested by Chang et al. (1988), is to calculate the estimates $\hat{\omega}_i$ and their respective statistics λ_i, where the subscript i makes reference to the outlier type, i=AO,IO,LS,TC,SO. The test statistic to use is

$$\eta(\tau, I) = \max_i \{|\lambda_i|\}. \tag{3.7}$$

If $\eta(\tau, I) > C$, where C is a predetermined critical value, then it is thought possible that the observed series is subject to the influence of an outlier of type i=I at time $t = \tau$.

The timing τ is seldom known a priori, but as shown in Chang et al. (1988), the likelihood ratio criterion leads to,

$$\eta_{T,J} = \max_t \eta(\tau_t, I_t) \qquad t = 1, \ldots, N, \tag{3.8}$$

where T denotes the period at which the maximum of $\eta(\tau_t, I_t)$ occurs and J the associated type of outlier. Then, if $\eta_{T,J} > C$, there is a possibility that the observed series is subject to the influence of one outlier of type J at time $t = T$. In order to compute the $\eta_{T,J}$ statistic above, the Kalman filter should be run on the vector of observations to obtain the vector $\mathbf{e}^*$, and on $N \times 5$ different $\mathbf{Z}_\mathbf{i}^*(\tau_\mathbf{t})$ vectors, for $i = AO, IO, LS, TC, SO$ and $t = 1, \ldots, N$. The procedure is thus computationally cumbersome. This problem can be overcome by using the filter $\Pi(B) = \phi(B)/\theta(B)$, appropiately truncated. The GLS estimator of ω can be computed by using the Kalman filter to obtain the vector of exact residuals $\mathbf{e}^*$, and then the truncated filter $\Pi(B)$ can be applied on $\mathbf{Z}^*$ to obtain the vector $\mathbf{Z}$.

Once the location and the type of the outlier are determined, its effect can be adjusted from the residuals using (3.4); the adjusted series can also be obtained as

$$y_t = y_t^* - \xi(B)\omega I_t^{(\tau)}. \tag{3.9}$$

In practice, the true parameters in α are usually unknown in the modelling stage, although they can be estimated consistently; the λ -statistic given by (3.6) still has in this case an asymptotic normal distribution (see Chang et al., 1988).

3.2 Multiple outliers

In a more general framework, one can consider the observed series as being affected by k deterministic shocks at times $t = \tau_1, \ldots, \tau_k$. In this case, the representation of y_t^* consists of,

$$y_t^* = \sum_{j=1}^{k} Z_{j,t}^*(\tau_j)\omega_j + y_t, \tag{3.10}$$

where $Z_{j,t}^*(\tau_j) = \xi_j(B) I_t^{(\tau_j)}$ represents the effect of the outlier at time $t = \tau_j$. The extended iterative procedure described in the Appendix does not detect the k outliers at the same time but proceeds in several iterations detecting them one by one. In the detection stage, the procedure starts by applying the Kalman filter on the vector of observations to obtain the residuals and the truncated filter $\Pi(B)$ on the vectors $Z_{j,t}^*(j)$ to determine the location and type of the k outliers in (3.10). Following Chen and Liu (1993), once the detection stage is completed, in order to avoid possible masking effects, the final ω_js are obtained within the following multiple regression model,

$$\mathbf{y}^* = \mathbf{Z}^*\omega + \mathbf{y},$$

where $\mathbf{Z}^*$ is an $N \times k$ matrix with columns $\mathbf{Z}_\mathbf{j}^*(\tau_\mathbf{j}) = (Z_{j,1}^*(\tau_j), \ldots, Z_{j,N}^*(\tau_j))$ and ω is a $k \times 1$ vector with elements ω_j. The application of the Kalman filter recursions on the vector of observations $\mathbf{y}^*$ and on the k columns of the matrix $\mathbf{Z}^*$ allows the specification of an OLS model, from which the vector ω can be estimated as in (3.5). Kohn and Ansley (1985) proposed an efficient way to estimate the vector ω using the QR algorithm, in which an orthogonal $N \times k$ matrix $\mathbf{Q}$ is obtained, such that $\mathbf{Q}'\mathbf{L}^{-1}\mathbf{Z}^* = (\mathbf{R}', \mathbf{0})'$, where $\mathbf{R}$ is a non-singular $k \times k$ upper-triangular matrix. Then $\hat{\omega} = \mathbf{R}^{-1}\mathbf{v}_1$, where $\mathbf{v}_1$ consists of the first k elements of the vector $\mathbf{v} = \mathbf{QL}^{-1}\mathbf{y}^*$.

Once $\hat{\omega}$ is obtained, residuals are identified and corrected, the linear series obtained, a new estimator of α computed, and iterations proceed as described in the Appendix.

4 Performance of the extended procedure

In Section 2 we presented three basic specifications for a seasonal outlier. Without any prior information, it would seem that requiring full cancellation of the effect in the next month is more restrictive than allowing for cancellation over the rest of the year. Moreover, while specification (2.5) can be exactly duplicated with two specifications of the type (2.6) or (2.7) of equal magnitude and applied, with opposite signs, to two consecutive months, no simple combination of specifications of the type (2.5) can reproduce the specifications (2.6) or (2.7). Thus (2.6) or (2.7) are more attractive candidates than (2.5). As we have seen, it is unlikely that, for many applications, empirical evidence could clearly discriminate between (2.6) or (2.7). We select specification (2.6), that is the SLS outlier given by

$$y_t^* = \frac{1}{1-B^s}\omega I_t^{(\tau)} + y_t, \tag{4.1}$$

simply because of its extreme simplicity.

In this section, we investigate the performance of the standard procedure when it is extended to include the SLS outlier type. We focus our analysis on two aspects of the procedure: i) the relative frequency of detection of at least one outlier while no one is effectively present, which is a measure of a type I error; and ii) the relative frequency of correct detection, which is a measure of the power. The simulations in this section were performed using the monthly "Airline model", popularized by Box and Jenkins (1970), and found to be often appropriate for series displaying trend and seasonality (see the large-scale study in Fischer and Planas, 2000). The model is given by equation (2.4) with a_t being a white-noise innovation, $-1 < \theta_1 < 1$, and $-1 < \theta_{12} < 0$. We performed a simulation in MATLAB, whereby, for each of the considered samples sizes (N=50,100, 200 or 400), 1000 noncontaminated "Airline" series were generated with $\theta_1 = \theta_{12} = -0.6$; the extended procedure, described in the Appendix, was then applied using critical values C=3.0,3.5,4.0 and 4.5. Table 4 gives the mean relative frequency of a type I error, decomposed into the different outlier types, for each combination of sample size and critical value.

		Outlier type						Outlier type				
		AO	IO	LS	TC	SLS		AO	IO	LS	TC	SLS
N=50	C=3.0	.05	.05	.07	.06	.08	C=4.0	.00	.00	.00	.00	.00
100		.13	.10	.11	.09	.16		.00	.00	.00	.00	.00
200		.16	.16	.17	.15	.22		.01	.01	.01	.00	.01
400		.18	.17	.21	.18	.24		.02	.03	.02	.01	.02
N=50	C=3.5	.01	.00	.01	.01	.01	C=4.5	.00	.00	.00	.00	.00
100		.03	.02	.02	.02	.01		.00	.00	.00	.00	.00
200		.06	.04	.06	.06	.06		.00	.00	.00	.00	.00
400		.11	.08	.10	.11	.12		.00	.00	.00	.00	.00

Table 4. Type I error in the extended procedure

As expected, the relative frequency of a type I error is a decreasing function of the critical value, but an increasing function of the number of observations. When the critical value is too low for the sample size, an SLS outlier is spuriously detected with a mean relative frequency slightly higher than for the other outlier types. However, when the critical value is adequate for the sample size ($C \geq 3.5$ for N=50, 100 and $C \geq 4.0$ for N=200,400), the mean relative frequency of a type I error is, for all cases, smaller than 5% and there are not significant differences among the outlier types. In a second simulation exercise, the influence of the parameters θ_1 and θ_{12} on the type I error was investigated. We considered values of $\theta_1 = 0.5, 0$ and -0.5 and $\theta_{12} = -0.1, -0.3, -0.5, -0.7$ and -0.9. As before, 1000 series of 100 observations each were generated. The results for this simulation exercise indicated that the relative frequency of a type I error is (almost) insensitive to changes in the parameters θ_1 and θ_{12}. (These results are not reported here but are available from the authors.)

Next, to investigate the power of the extended iterative procedure in terms of outlier detection, we study the relative frequency of correct detection (type and location are correctly identified) of one outlier. The "Airline" model with $\theta_1 = -0.6$ and $\theta_{12} = -0.1, -0.3, -0.5, -0.7, -0.9$ is used to generate simulated series of 100 observations. We consider one outlier affecting the observed series at the middle of the sample, $t = 50$. The size of the initial impact ω is considered equal to 4 or 5. For each combination of outlier types and parameters θ_{12}, 1000 simulated series were generated. Table 5 reports the mean relative frequency of correct detection for critical values C=3.5 and C=4.0.

		Outlier type									
		AO		TC		LS		IO		SLS	
		C=3.5	4.0	3.5	4.0	3.5	4.0	3.5	4.0	3.5	4.0
θ_{12}=-.1	ω=4	.97	.93	.96	.94	.99	.99	.51	.32	.67	.50
	5	1	1	.99	.99	1	1	.79	.66	.88	.88
-.3	4	.95	.85	.92	.85	.98	.95	.50	.33	.74	.60
	5	.99	.98	.98	.98	.99	.99	.79	.66	.92	.91
-.5	4	.87	.77	.87	.73	.95	.90	.50	.32	.90	.79
	5	.98	.97	.97	.95	.99	.99	.79	.66	.98	.98
-.7	4	.79	.62	.77	.65	.91	.84	.49	.31	.99	.97
	5	.95	.90	.93	.93	.99	.98	.78	.65	1	1
-.9	4	.69	.54	.71	.52	.87	.74	.44	.27	1	1
	5	.90	.86	.90	.84	.97	.95	.71	.58	1	1

Table 5. Mean relative frequency of correct detection using the extended procedure

The power of the extended procedure increases with the size of the initial impact and decreases with the critical value. Table 6 shows that the power is a decreasing function of the absolute value of the parameter θ_{12} for all outlier types except for the SLS. The AO and TC outliers are seen to perform equally well and the LS outlier outperforms both. When seasonality is highly moving, the SLS outlier performs worse than the previous three, although when seasonality is stable, it offers the best performance. For all cases, the IO presents, by far, the worst results in terms of power of the procedure (always less than 80% and, in 7 out of 20 cases, less than 50%).

5 Consequences of not including the SLS type

In this section, we compare the performance of the standard iterative procedure with that of the extended procedure when the observed series is subject to the presence of one SLS.

Using the monthly Airline model in (2.4) with $\theta_1 = -0.6$ and θ_{12}=-.1, -.3, -.5, -.7 and -.9, 1000 series of 100 observations each were generated for each combination of θ_1 and θ_{12}. Next, the series were contaminated at time $t = 50$ with one SLS of size $\omega = 4$ or $\omega = 5$, and the standard iterative procedure was applied three times with critical values C=3.5, C=4.0 and C=4.5. Table 6 presents the mean relative frequency of wrongly detecting one or more IO (columns labeled IO) or of not detecting any outlier (columns labeled NO). Notice that the relative frequency of detecting any of the other three outlier types different from an IO is extremely low.

		θ_{12}									
		-.1		-.3		-.5		-.7		-.9	
C	ω	NO	IO	NO	IO	NO	IO	NO	IO	NO	IO
3.5	4	.37	.56	.41	.58	.50	.50	.55	.45	.73	.26
	5	.11	.83	.14	.84	.17	.82	.28	.72	.53	.46
4.0	4	.65	.32	.68	.31	.69	.31	.80	.20	.90	.10
	5	.30	.67	.33	.66	.40	.60	.51	.49	.81	.19
4.5	4	.86	.13	.88	.12	.90	.10	.92	.08	.98	.02
	5	.56	.42	.58	.41	.67	.33	.79	.21	.97	.03

Table 6. Performance of the standard procedure for a SLS-contaminated series

The results in Table 6 indicate that, when the time series is subject to the presence of one SLS, the standard outlier detection procedure performs poorly. The mean relative frequency of wrongly detecting one IO is high, being larger than 50% in 9 out of 30 cases; this frequency increases with the size of the initial impact, ω. Table 6 shows a trade-off between not detecting any outlier and detecting one IO. This trade-off depends on the value of the parameter θ_{12} and, therefore, on the stability or instability of the seasonal component. When the seasonal component is highly erratic, it is more likely to detect one

IO than nothing and, in contrast, when the seasonal component is very stable, the procedure, most often will not detect any outlier.

We now compare the estimates of the parameter θ_{12} when the "cleaning" or preadjustment of the observed series is carried out using the standard procedure or the extended procedures. For this comparison, we generated 1000 Airline series of 100 observations each, with $\theta_1 = -0.6$ and $\theta_{12} = -0.6$, and then included one SLS of size $\omega = 5$ at time t=50. Table 7 presents the first two moments of the distribution for $\hat{\theta}_{12}$ when the SLS is correctly adjusted (using the extended procedure) and, when it is treated as an IO or not adjusted (using the standard procedure).

	θ_{12}		σ_a	
	Mean	Std. Error	Mean	Std. Error
True parameter	-.600	-	1.000	-
Standard procedure	-.430	.097	1.081	.096
Extended procedure	-.626	.125	.987	.082

Table 7. θ_{12} and σ_a estimates for a SLS contaminated series

In Figure 7 the two histograms are displayed and approximated by the normal distribution. and it is seen that the density for the standard procedure is strongly biased. The conclusion is, therefore, that ignoring or missidentifying one SLS leads to an important bias in the estimation of the parameter θ_{12} towards the region of unstable seasonality, which would have the effect of yielding estimates of the seasonal component unreasonably erratic.

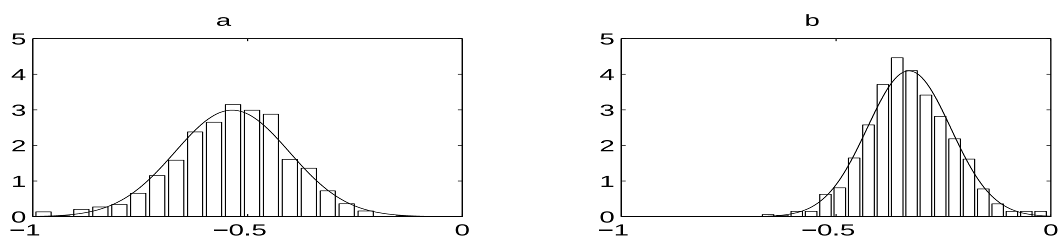

Figure 7. Densities for θ_{12} estimates. a) Histogram and normal approximation for the Extended procedure; b) Histogram and normal approximation for the Standard procedure.

Finally, we investigate the consequences of ignoring or missidentifying one SLS on some diagnostics for the model. In particular, we consider the effects on the residual autocorrelation ρ_{12} and on the Box-Ljung statistic Q(36) for the residuals. Using the SLS-outlier contaminated series that were generated for the simulation exercise of Table 7, the standard and extended iterative procedures were applied with critical value C=4.0. The values of $\hat{\rho}_{12}$ and

Q(36) obtained for the residuals with the two procedures are compared in Figures 8 and 9. Each point in Figure 8 is a representation of the pair of values $(\hat{\rho}_{12}^{SLS},\ \hat{\rho}_{12}^{IO})$ and each point in Figure 9 is a representation of the pair of values $(Q(36)^{SLS},\ Q(36)^{IO})$. Points on the straight line (x=y) indicate cases for which the two procedures lead to the same value of $\hat{\rho}_{12}$ or Q(36); values over the line indicate cases for which the standard procedure leads to values of $\hat{\rho}_{12}$ or Q(36) larger than for the extended procedure. Figures 8 and 9 indicate that, as seasonality becomes more stable, the ρ_{12} and Q estimates obtained with the two procedures tend to diverge. The use of the standard iterative procedure, when an SLS is present on the observed series, may lead to significant $\hat{\rho}_{12}$ coefficients for the residuals, indicating that not all seasonality has been removed, and to high values of the Ljung-Box statistics, indicating the presence of autocorrelation in the residuals.

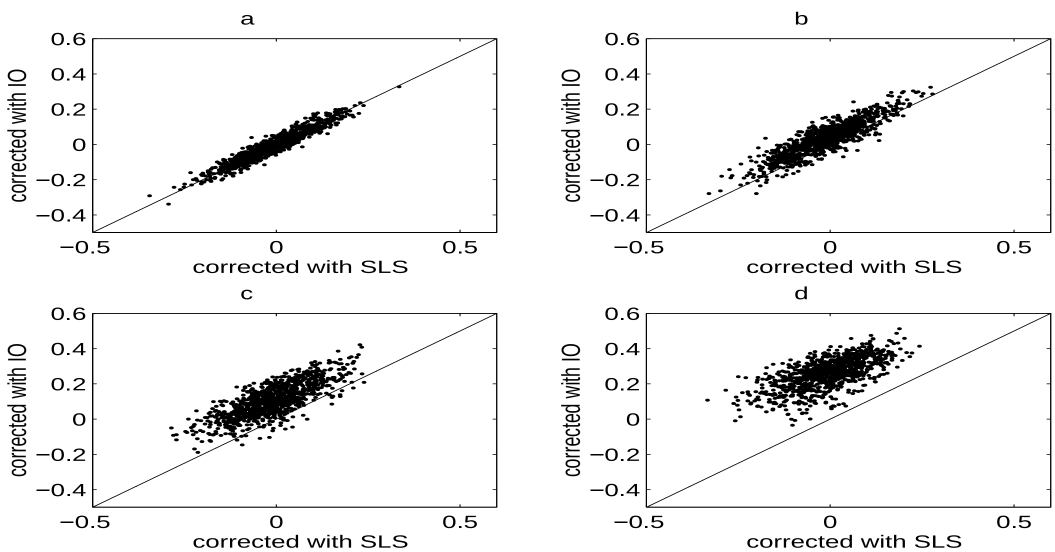

Figure 8. Coefficient ρ_{12} for residuals in an Airline model with $\theta_1 = -0.6$ and a) $\theta_{12} = -0.1$; b) $\theta_{12} = -0.5$; c) $\theta_{12} = -0.7$; d) $\theta_{12} = -0.9$.

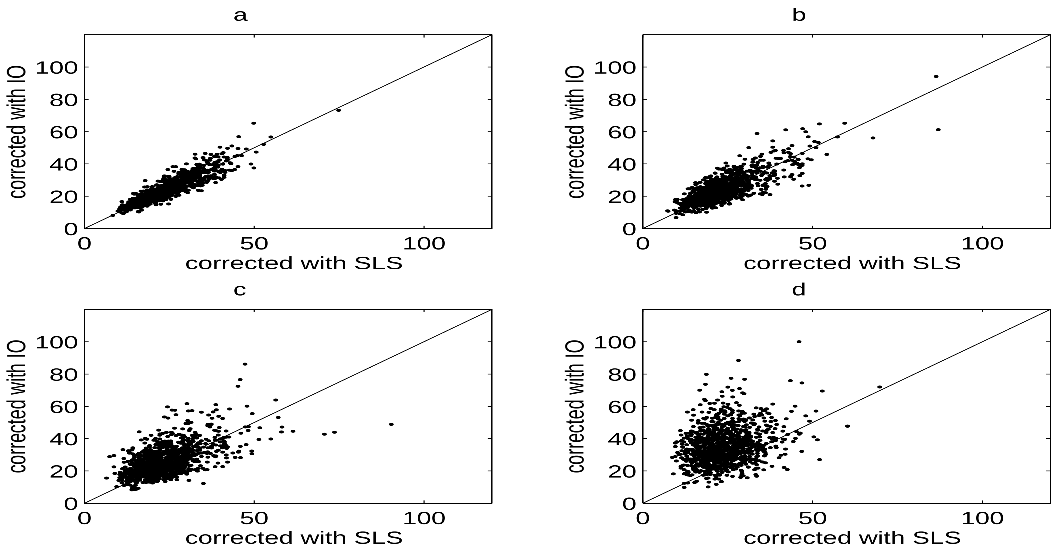

Figure 9. Box-Ljung statistic for residuals in the Airline model with $\theta_1 = -0.6$ and a) $\theta_{12} = -0.1$; b) $\theta_{12} = -0.5$; c) $\theta_{12} = -0.7$; d) $\theta_{12} = -0.9$.

6 Some examples

We consider four real monthly series as examples. In all of them the automatic outlier detection procedure of TRAMO, that uses the standard four types of outliers (the standard procedure), detected innovational outliers. Two of the series are production series, one of them, an aggregate index, namely, the Italian General Industrial Production Index (IIPI) of Section 2, and the other a component of a general index, namely, the Spanish Production of Metal Products (SPMP). The third series is an aggregate price index, namely the European Industrial Production Price Index (EPPI), and the fourth example is an inventory series, namely, the U.S. South Manufactured Homes Inventories (SMHI). The series span the periods 1981.1-1996.12, 1981-1993.12, 1980.1-1997.7, and 1980.1-1998.3, respectively, and are displayed in Figure 10.

The automatic model identification and automatic outlier detection and correction procedure of TRAMO was applied. The models identified for the IIPI and the SPMP series were Airline-type models as in (2.4), with 2 and 5 outliers respectively, and significant trading day and Easter effects in both cases. For the EPPI series a $(0, 2, 1)(0, 1, 1)_{12}$ model with one outlier is selected, and for the SMHI series a purely regular (1,2,1) model with no seasonality and 5 outliers (two of them IOs) is chosen. The Ljung-Box Q-test (always computed with 24 residual autocorrelations) for the IIPI series is 38.0, larger than the critical value $\chi^2_{22}(.05) = 33.9$. For the SPMP and the EPPI series, the

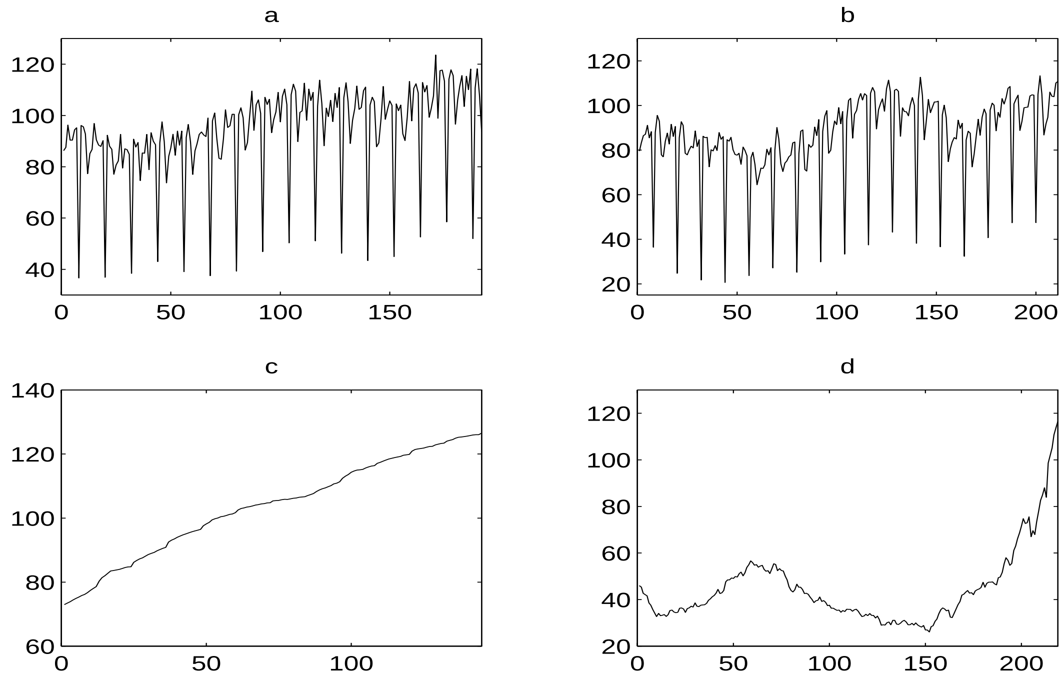

Figure 10. a) Italian industrial production index b) Spanish production of metal products c) European industrial production price index d) South manufactured homes inventories

Bowman-Shenton test for normality of the residuals yields the values 7.0 and 6.9, respectively, both above the critical value $\chi^2_2(.05) = 5.99$. In order to see whether undetected outliers could be responsible for the failure of these tests, the critical value for outlier detection in the automatic procedure was lowered from 3.5 to 3.3 for the three series. The outliers obtained are given in Table 8 and all three series are seen to contain some IOs. Further, for the two production series, outliers pile up for August, while for the price series they are concentrated in January. As shown in Table 9, the tests are now comfortably passed. For the SMHI series, the residual autocorrelations for lag 12, 24, and 36 are all positive (with $\rho_{12} = .22$) and the first autocorrelations show a clear annual periodicity. Besides, the Pierce Q_s-statistics for seasonal autocorrelation in the residuals is equal to 12.2, larger than the critical value $\chi^2_2(.05) = 5.99$. Further, the residuals display some heteroscedasticity and their mean seems to oscillate in time. The model specification was changed to that of an Airline model in the levels. The results were not acceptable: the Q-statistic becomes 56.7 and the first 21 residual autocorrelations are positive. Given that the mean is significant and $\hat{\theta}_1 > 0$, an obvious modification is to increase the number of regular differences to 2, and to protect against overdifferencing by increasing the order of the regular MA also to 2. The new model (in the levels and without mean) contains the 5 outliers given in Table

8, concentrated towards the end of the sample period and occurring at the beginning of a year or in the early Fall. As shown in Table 9, the tests are passed with no problem.

Series	Procedure	C	Type	Date of Detected outliers
IIPI	Standard	3.3	AO	1984.8, 1995.8, 1987.1,1990.8,1984.4
			TC	1989.8,1992.12
			IO	1994.8
	Extended	3.5	AO	1984.8, 1987.1
			SLS	1994.8, 1988.8
EPPI	Standard	3.3	IO	1986.1, 1985.1, 1987.1
			LS	1988.1
	Extended	3.5	SLS	1985.12, 1986.12
SPMP	Standard	3.3	AO	1980.8, 1983.8, 1990.8,1982.8
			TC	1982.12
			IO	1985.8, 1994.8
	Extended	3.5	AO	1990.8
			SLS	1981.8, 1984.8, 1994.8
SHMI	Standard	3.5	AO	1997.9, 1997.1, 1995.1
			IO	1996.4, 1996.10
	Extended	3.5	AO	1997.9
			LS	1997.1
			TC	1996.2
			SLS	1996.3, 1996.9

Table 8. Detected outliers using the standard and the extended procedures

Series	Procedure	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_{12}$	BIC	σ_a	Q(24)	Q_s	N	outliers
IIPI	Standard C=3.3	-.623	-	-.655	-7.22	.024	21.08	3.49	0.36	8
	Extended C=3.5	-.602	-	-.781	-7.22	.024	29.58	3.49	0.33	4
EPPI	Standard C=3.3	-.462	-	-.594	-12.75	.001	16.72	2.03	1.64	4
	Extended C=3.5	-.339	-	-.729	-12.81	.001	13.76	0.73	1.34	2
SPMP	Standard C=3.3	-.541	-	-.406	-6.06	.043	21.04	1.44	4.03	7
	Extended C=3.5	-.503	-	-.654	-6.09	.043	25.09	1.26	4.42	4
SHMI	Standard C=3.5	-.687	-.140	-.739	.769	1.369	16.21	3.87	1.53	5
	Extended C=3.5	-.832	-	-.959	.533	1.213	23.61	3.62	2.13	5

Table 9. Summary of ARIMA estimation results.

The extended procedure, including the SLS outlier type, was then applied with the default value for the critical level (C=3.5). The location and type of the detected outliers can also be found in Table 8. It is seen that the total number of outliers has decreased from an average of 6 outliers per series (a slightly large number, but, considering the series length, by no means exorbitant) to an average of 4 outliers per series (or, approximately, 2 outliers per 100 observations). In the four cases, SLS outliers are detected; on the contrary, no IO is now present. The effect of the SLS outliers is represented in Figure 11. The results for the estimation of the ARIMA models are summarized in Table 9 and there is no evidence of misspecification in any of the new models. For the EPPI, SPMP, and SMHI series, the BIC criterion indicates preference for the model with SLS outliers; for the IIPI series, it does not distinguish between the two models. The θ_{12} estimates are closer to -1 for the extended procedure, indicating more stable seasonal components. This effect can be seen in Figure 12 which compares the stochastic seasonal component obtained with the standard and the extended procedures: the later has removed heterocedasticity from the seasonal component, which behaves now in a more regular manner. (This result is also obtained in Kaiser and Maravall (2000) using an outlier specification consisting of a SLS applied to two consecutive months.)

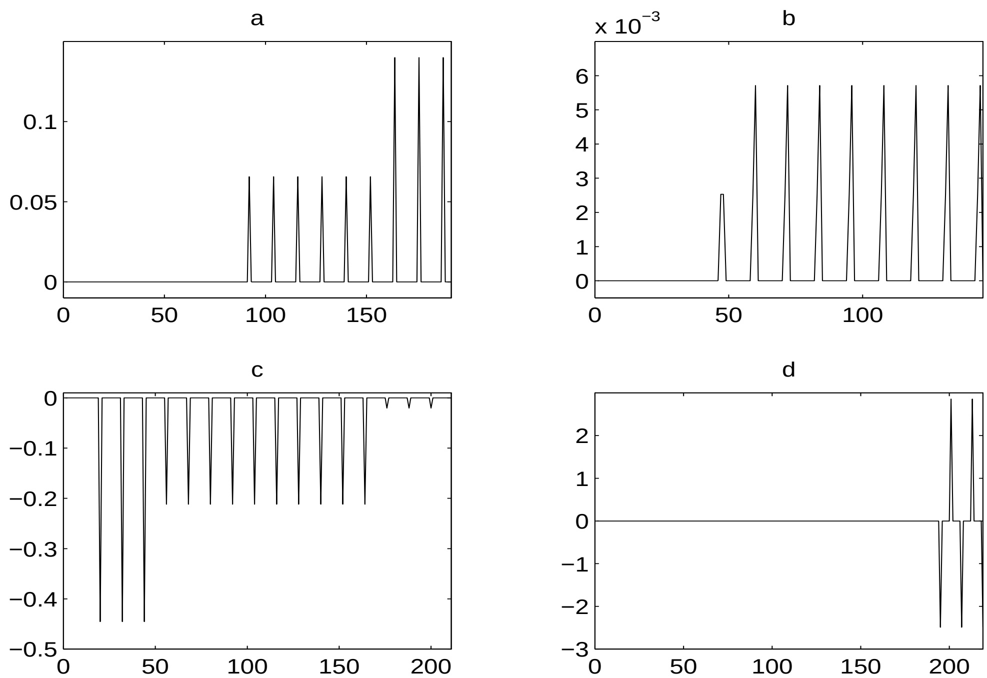

Figure 11. Effect of detected SLS outliers on the observed series. a) IIPI series b) EPPI series c)SPMP series d)SMHI series

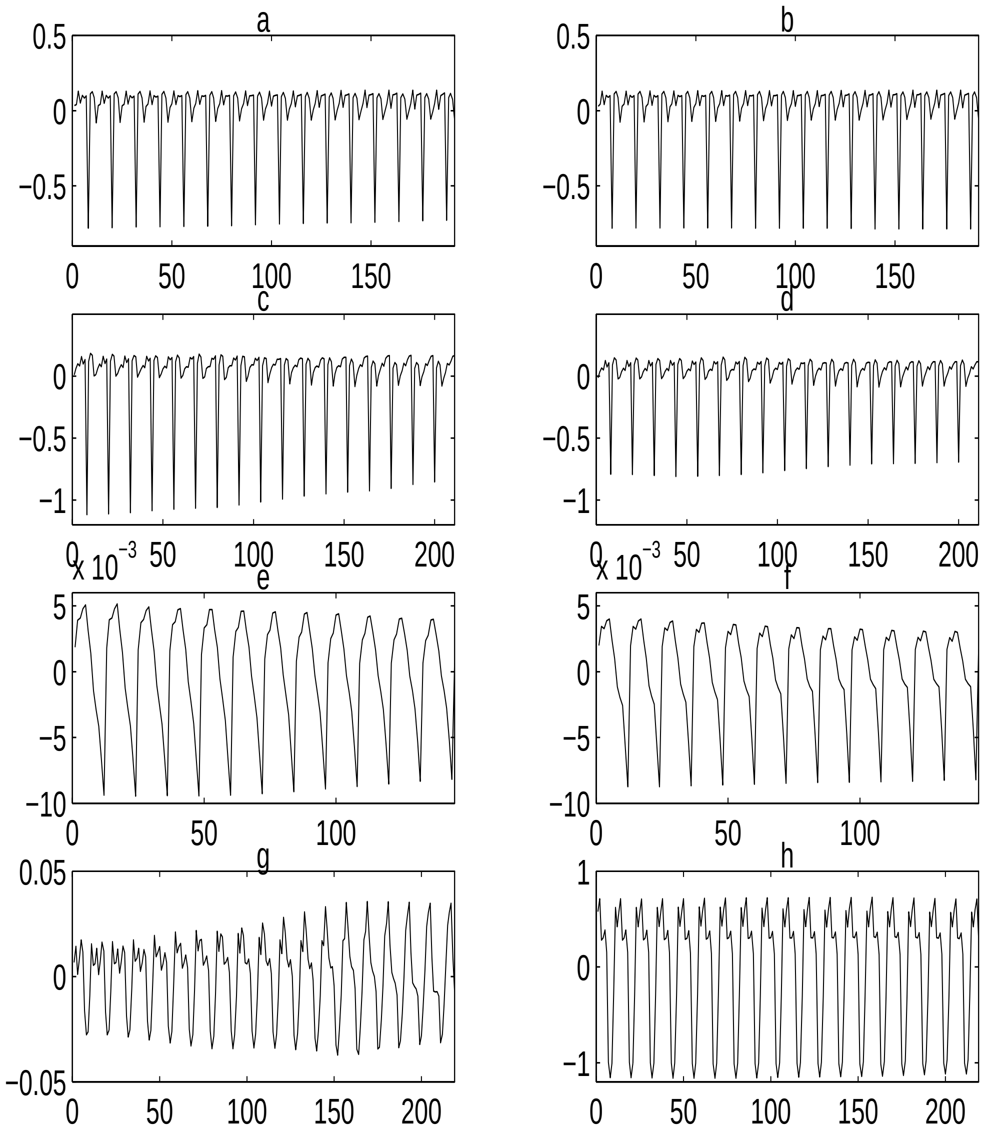

Figure 12. Stochastic Seasonal component. IIPI series a) standard procedure b) extended procedure; SPMP series c) standard procedure d) extended procedure; EPPI series e) standard procedure f) extended procedure; SHMI series g) standard procedure h)extended procedure

To further compare the models obtained with the two procedures, two out-of-sample forecasting exercises were performed. First, the models were re-estimated with 20 fewer observations (C=3.3 for the standard procedure; C=3.5 for the extended one). One-period-ahead forecasts were then computed for the last 20 months. (The SHMI series was not included in the exercise because all outliers occur at the end of the series and fall in the out-of-sample period). Table 10 gives the out-of-sample variance and the value of the F-test for the equality of the variances (approximate critical value 1.6). The comparison of these magnitudes reveals that the models that include SLS outliers have smaller out-sample variance and smaller values for the F-test.

	IIPI		EPPI		SPMP	
	Standard	Extended	Standard	Extended	Standard	Extended
Variance	.1051E-02	.4780 E-03	.1475 E-05	.1189 E-05	.2504 E-02	.1185 E-02
F-test	1.817	.805	.632	.483	1.545	0.618

Table 10. Summary of the out-of-sample forecasting results.

Finally, to asses the stability of the models, the following exercise was performed (as before, the SMHI series was not considered). The series GIPI (192 observations) was stopped at observation 92. The model was reestimated, including the outliers detected for that subperiod and, with all parameters fixed, one-period-ahead forecasts were subsequently computed for the last 100 observations. A similar computation was made for the EUPRIN series (144 observations), reestimating in this case, the model for the first 64 observations, and obtaining the one-period-ahead forecast errors for the next 80 observations. Finally, for the METIPI series (211 observations), the model for the first 121 observations was reestimated, and one-period-ahead forecast errors were obtained for the next 90 observations. Table 11 compares the forecast standard errors. From the table, it is seen that the extended method decreases, in the three cases, the forecast standard error, although the differences are small. Comparing the forecast standard errors for the out-of-sample period with the in-sample residual standard error, given in Table 9, the proximity of the two is remarkable. In summary, the in-sample behavior of the model is clearly stable.

Series	Procedure	Forecast SE
IIPI	Standard	.0288
	Extended	.0276
EPPI	Standard	.0018
	Extended	.0015
SPMP	Standard	.0139
	Extended	.0125

Table 11. Standard error of the one-period-ahead forecast.

7 Conclusion: innovational versus seasonal outlier

In Section 2 we mentioned that IOs have some important undesirable features. In particular, for series containing trend and seasonality, such as the ones that follow, for example, a model of the (often found) type

$$\nabla\nabla_s y_t = \theta(B) a_t, \tag{7.1}$$

where $\theta(B)a_t$ is a stationary process, the IO has the form

$$IO_t = \omega \frac{\theta(B)}{\nabla\nabla_s} I_t^{(\tau)},$$

and as Figures 1d and 2 illustrated, its effect becomes larger and larger. An IO imposes thus a relatively complicated and quantitatively important deterministic effect on the series, which eventually may account for a large share of the series level.

We further showed in Section 4 that the test used to detect outliers has, for the case of the IO, consistently low power for different model specifications, considerably lower than for the other types of outliers (AO, LS, TC or SLS)

Considering the examples of Section 6, in all cases the automatic procedure that considers the standard four types of outliers yields innovational outliers. The extended procedure (which simply adds the SLS outlier to the standard method) provides a clear improvement in results, with a more parsimonious outlier representation. Moreover, all IOs disappear and SLS outliers show up in their place. All considered, a conclusion emerges: instead of adding an additional outlier, the automatic procedure could be simply modified by replacing the IO outlier with the SLS outlier given by (2.6), or its purely seasonal component, given by (2.7).

References

BALKE, N.S. (1993), "Detecting Level Shifts in Time Series". *Journal of Business and Economic Statistics*, 11, 81-92

BELL,W.R. (1983), "A Computer Program for Detecting Outliers in Time Series", *Proceedings of a Conference of the Business and Economic Statistics Section of ASA.*

BELL, W.R. AND HILLMER, S.C. (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models", *Journal of Time Series Analysis*, 12, 283-300.

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

Box, G.E.P., Hillmer, S.C. and Tiao, G.C. (1978), "Analysis and Modeling of Seasonal Time Series", in Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series,* Washington, D.C.: U.S. Dept. of Commerce. Bureau of the Census, 309-334.

Box, G.E.P. and Tiao, G.C. (1975), " Intervention Analysis with Applications to Economic and Enviromental Problems", *Journal of the American Statistical Association* 70, 71-79.

Chang, I. (1982), " Outliers in Time Series". Unpublished Ph.D. thesis, University of Wisconsin-Madison, dept. of Statistics.

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

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

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen, B.C. (1998), " New Capabilities and Methods of the X12 ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 12, 127-177.

Fischer, B. and Planas, C. (2000) "Large Scale Fitting of Regression Models with ARIMA Errors, *Journal of Official Statistics, 16, 173-184.*

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

Gómez, V. and Maravall, A. (2001), "Automatic modelling methods for univariate series", Ch.7 in Peña, D., Tiao, G.C., and Tsay, R.S. (eds.), *A Course in Advanced Time Series Analysis*, New York: Wiley.

Gómez, V., and Maravall, A. (1996), "Programs TRAMO and SEATS, Instructions for the user (Beta version: September 1996)", Working Paper 9628, Research Department, Bank of Spain.

Gómez, V., and Maravall, A. (1994), " Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter", *Journal of the American Statistical Association*, 89, 611-624.

HARVEY, A.C. (1989), *Forecasting Economic Time Series and the Kalman Filter*, Cambridge: Cambridge University Press.

JONES, R. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations", *Technometrics*, 22,389-395.

KAISER, R. (1999), "Detection and Estimation of Structural Changes and Outliers in Unobserved Components", *Computational Statistics*, 14, 533-558.

KAISER, R. AND MARAVALL, A (2000), "An application of TRAMO-SEATS: Changes in Seasonality and Current Trend-Cycle Assessment", Working Paper 0011, Research Department, Bank of Spain; to appear in *Proceedings of the IV International Conference on Establishment Surveys* (invited papers).

LIU, L.M., HUDAK, G., BOX, G.E.P., MULLER, M.E., AND TIAO, G.C. (1986), *The SCA Statistical System: Reference Manual for Forecasting and Time Series Analysis*, Dekalb, IL: Scientific computing Associates.

KOHN, R., AND ANSLEY, C.F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models", *Biometrika*, 72, 694-697.

PEÑA, D.(2001), "Outliers, Influential Observations, and Missing Data", Ch.6 in Peña, D., Tiao, G.C., and Tsay, R.S. (eds.), *A Course in Advanced Time Series Analysis*, New York: Wiley.

PEÑA, D. (1990), "Influential Observations in Time Series", *Journal of Business and Economic Statistics* 8, 235-241.

PROIETTI, T. (1998), "Seasonal Heteroscedasticity and Trends", *Journal of Forecasting*, 17, 1-17.

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

A extended procedure for outlier detection

The procedure starts with the specification of the model for the observed series as if there were no outliers, and the following stages are followed:

I.1 Obtain the maximum likelihood estimators for the unknown parameters in vector α based on the vector of standardized residuals $\mathbf{e}^*$. In the first iteration, the residuals obtained from the application of the Kalman filter on the observed series are used; after the first iteration, the adjusted residuals are used to evaluate the likelihood function.

Detection inner loop

I.2 For $t = 1, \ldots, N$ and $i = AO, IO, LS, TC, SO$, compute $\lambda_i(t)$ using (3.6), and the statistic $\eta(t) = max_i \{|\lambda_i(t)|\}$. If $\eta_\tau = max_t \eta(t) > C$, where C is a predetermined critical value, then there is a possibility of one outlier at time $t = \tau$.

In the presence of outlying observations, the standardized residuals, obtained with the Kalman filter, are contaminated and, hence, $\hat{\sigma}_a^2 = \{1/N\}\mathbf{e}'^*\mathbf{e}^*$ may be overestimated. One method to overcome this problem, which is not time consuming, is the omit-one method, in which $\hat{\sigma}_a^2$ is calculated, such that the residual at time $t = \tau$ is omitted. Other alternatives include the MAD method or the α % trimmed method, see Chen and Liu (1993).

I.3 If a possible outlier is found, remove its effect from the residuals and obtain the adjusted residuals $\mathbf{e}$ using,

$$\mathbf{e} = \mathbf{e}^* - \mathbf{Z}\hat{\omega},$$

and go back to I.2 to iterate. Otherwise, proceed to I.4.

I.4 If no outlier is found in the first iteration, then stop. If one or more outliers have been detected in the previous iterations from steps I.2-I.3, then go back to I.1 to revise the parameter estimates. Continue to repeat I.1-I.3 until no new outliers are found, then go to II.

Joint estimation stage

II The effects of the identified outliers are jointly estimated in the multiple regression model in (3.10) by applying the Kalman filter on the vector of observations and on each column in matrix $\mathbf{Z}$; and applying the QR algorithm on $\mathbf{Z}$. This provides an efficient estimator for vector ω. Compute the t-statistic for the estimated effects and check if there is any outlier for which the t-statistic is smaller than C, where C is the critical value used in I.2. If there are not, then obtain the adjusted series, check whether the initial specification of its ARIMA model is still valid, apply the Kalman filter on the adjusted series, obtain the new residuals and go back to stage I to repeat the complete process until no new outliers can be detected. Otherwise, delete one by one the insignificant effects and re-estimate the multiple regression model until all the ω_is are significant, then obtain the vector of adjusted observations, apply the Kalman filter on it, obtain the new residuals and go back to stage I to iterate until no new outliers can be detected.