13

Additive Outlier Detection in Seasonal ARIMA Models by a Modified Bayesian Information Criterion

Pedro Galeano and Daniel Peña

CONTENTS

13.1 Introduction

The detection of outliers in a time series is an important issue because their presence may have serious effects on the analysis in many different ways. For instance, even if the time series model is well specified, outliers can lead to biased parameter estimation, which may result in poor forecasts. Several outlier detection procedures have been proposed for detecting different outlier types in autoregressive-integrated-moving average (ARIMA) time series models, including those proposed in Fox (1972), Tsay (1986, 1988), Chang et al. (1988), Chen and Liu (1993), McCulloch and Tsay (1994), Luceño (1998), Justel et al. (2001), Bianco et al. (2001), and Sánchez and Peña (2003), among others. Most of these methods are based on sequential detection procedures that first search for the presence of an outlier. When the first outlier is found, its size is estimated, its effect is cleaned from the series, and a new search for outliers is carried out. However, as Sánchez and Peña (2003) pointed out, sequential detection procedures have three main drawbacks. First, a biased estimation of the initial parameter values may strongly affect the power to detect the outliers. Second, in many situations, the distributions of the test statistics are unknown and critical values needed to apply the tests should be estimated via simulation for different sample sizes and models. Indeed, iterative procedures sequentially test for the presence of outliers, which usually leads to overdetection of their number as a consequence of the inability to control the size of the outlier tests. Third, they suffer from both the masking effect, which means that outliers are undetected because of the presence of others, and the swamping effect, which means that outliers affect the data in such a way that good observations appear to be outliers as well.

The main purpose of this chapter is to develop a procedure for detecting additive outliers in seasonal ARIMA time series models based on model selection strategies. The proposed procedure is designed to try to mitigate the drawbacks of sequential detection methods. In order to achieve this goal, it is shown that the problem of detecting additive outliers in seasonal ARIMA models can be formulated as a model selection problem in which the candidate models explicitly assume the presence of additive outliers at given time points. Therefore, the problem of detecting additive outliers reduces to the problem of selecting the best model, which is the one that contains the true outliers in the series. It is important to note that this chapter is focused on the detection of additive outliers, which are especially pernicious, for instance, in unit root testing; see Perron and Rodriguez (2003). Although the proposed methodology can be extended to additional types of outliers, this requires more elaboration and is beyond the scope of this chapter. Therefore, this chapter can be seen as a first attempt at outlier detection in time series based on model selection strategies.

Model selection is one of the most important problems in statistics and consists in selecting, from a set of candidate models, the one that best fits the data under some specific criteria. Two main strategies have been developed: the goal of the efficient criteria is to select the model that it is expected to best predict new observations, while the goal of the consistent criteria is to select the model that actually has generated the data. These strategies lead to different model selection criteria. The efficient criteria include, among others, the Final Prediction Error (FPE), proposed by Akaike (1969), which is an estimator of the one-step-ahead prediction variance; the Akaike Information Criterion (AIC), proposed by Akaike (1973), which is an estimator of the expected Kullback–Leibler divergence between the true and the fitted model; and the corrected Akaike Information Criterion (AICc), derived by Hurvich and Tsai (1989), which is a bias-corrected form of the AIC that appears to work better in small samples. These criteria have the property that, under the main assumption that the data come from a model with an infinite number of parameters, they asymptotically select the model producing the least mean squared prediction error. The consistent criteria include, among others, the Bayesian information criterion (BIC), derived by Schwarz (1978), which approaches the posterior probabilities of the models; and the Hannan and Quinn

Criterion (HQC), derived by Hannan and Quinn (1979), which was designed to have the fastest convergence rate to the true model. These criteria have the property that, assuming that the data come from a model with a finite number of parameters, the criteria will asymptotically select the true model.

This chapter proposes a new model selection criterion for selecting the model for a time series that follows a seasonal ARIMA model and is contaminated by additive outliers. The proposed model selection criterion avoids the use of multiple hypothesis testing, iterative procedures, and the simulation of critical values. As the objective is to incorporate in the final model the true number of additive outliers in a time series, the model selection criterion considered in this chapter falls more naturally into the category of consistent criteria. Therefore, we explore some modification of the Bayesian information criterion including an additional term useful for outlier detection. However, computation of the values of the criterion for all the possible candidate models, including all the possible configurations of outliers, may be impossible even for small sample sizes. Therefore, this chapter also proposes a procedure for selecting the most promising models.

The remainder of this chapter is organized as follows. In Section 13.2, the additive outlier detection problem for seasonal ARIMA models is formulated as a model selection problem. Section 13.3 presents the modified Bayesian information criterion for these models. Section 13.4 proposes a procedure for selecting the most promising models. Finally, Section 13.5 is devoted to showing the performance of the procedure by means of simulated and real data examples.

13.2 Formulation of the Outlier Detection Problem

A time series x_t follows a seasonal ARIMA$(p,d,q) \times (P,D,Q)_s$ model if,

$$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d x_t = \Theta_Q(B^s)\theta_q(B)\epsilon_t, \qquad (13.2.1)$$

where B is the backshift operator such that $Bx_t = x_{t-1}$; $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are regular backshift operator polynomials of finite degrees p and q, respectively; $\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$ and $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \cdots - \Theta_Q B^{sQ}$ are seasonal backshift operator polynomials with seasonal period s of finite degrees P and Q, respectively; d is the number of regular differences; D is the number of seasonal differences; and ϵ_t is a sequence of independent and identically distributed Gaussian random variables with zero mean and standard deviation σ. It is assumed that the roots of $\phi_p(B)$, $\theta_q(B)$, $\Theta_Q(B^s)$, and $\Phi_P(B^s)$ are all outside the unit circle and that neither the polynomials $\phi_p(B)$ and $\theta_q(B)$ nor $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ have common factors. In the case of $D = P = Q = 0$, the model in (13.2.1) reduces to the nonseasonal ARIMA model.

Suppose now that the observed time series $(y_1, \ldots, y_T)$ contains m additive outliers. Therefore,

$$y_t = x_t + w_{t_1} I_t^{(t_1)} + \cdots + w_{t_m} I_t^{(t_m)},$$

where x_t follows the seasonal ARIMA$(p, d, q) \times (P, D, Q)_s$ model in (13.2.1), $\tau_m = (t_1, \ldots, t_m)'$ is the $m \times 1$ vector containing the locations of the outliers, for each $h \in \tau_m$, $I_t^{(h)}$ is a dummy variable such that $I_t^{(h)} = 1$ if $t = h$ and is zero otherwise, and $w_{t_1}, \ldots, w_{t_m}$ are the outlier sizes at the corresponding locations. Consequently, the time series y_t follows the regression model with seasonal ARIMA errors given by

$$\begin{aligned} \Phi_P(B^s)\phi_p(B)(1 - B^s)^D(1 - B)^d(y_t - w_{t_1} I_t^{(t_1)} - \cdots - w_{t_m} I_t^{(t_m)}) \\ = \Theta_Q(B^s)\theta_q(B)\epsilon_t, \end{aligned} \tag{13.2.2}$$

in which the regressors are the dummy variables and the parameters linked with the regressors are the outlier sizes. This model is denoted as M_{τ_m}. Note that this notation suppresses whichever combination $t_1, \ldots, t_m$ and seasonal ARIMA model are being considered, but this will be clear in the context. The parameters of the model M_{τ_m} can be summarized in the $p_m \times 1$ vector given by

$$\rho_{\tau_m} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \Phi_1, \ldots, \Phi_P, \Theta_1, \ldots, \Theta_Q, w_{t_1}, \ldots, w_{t_m}, \sigma)',$$

where $p_m = p + q + P + Q + m + 1$.

Let $\mathcal{L}(\rho_{\tau_m} | y, M_{\tau_m})$ be the likelihood function of the time series $y = (y_1, \ldots, y_T)'$, given the model M_{τ_m} and its parameters ρ_{τ_m}. Exact maximum likelihood estimates (MLEs) of the model parameters, denoted as $\widehat{\rho}_{\tau_m}$, are obtained after maximizing the likelihood $\mathcal{L}(\rho_{\tau_m} | y, M_{\tau_m})$ with respect to the parameters ρ_{τ_m}. Several methods for maximizing the likelihood of seasonal ARIMA regression models such as the one in (13.2.2) are available. See, for instance, the methods proposed by Harvey and Phillips (1979), Kohn and Ansley (1985), and Gómez and Maravall (1994), among others. In particular, in the simulated and real data examples in Section 13.5, the `arima` function implemented in the statistical software R (http://www.r-project.org/) is used. This function computes the exact likelihood via a state-space representation of the ARIMA process, and the innovations and their variances are found by a Kalman filter.

In summary, given the time series y, the number and location of the additive outliers, m and τ_m, respectively, and the parameter vector, ρ_{τ_m}, are unknown and have to be estimated from the observed time series. Determining the number and location of outliers in y is now equivalent to selecting the model M_{τ_m} with the true outliers from among the set of candidate models. Once this is done, inference on the vector of parameters, ρ_{τ_m}, can be carried out by means of the MLEs, $\widehat{\rho}_{\tau_m}$. However, note that such inferences are conditional on the assumption that the true outliers have been selected.

13.3 Modified Bayesian Information Criterion for Outlier Detection

Once the outlier detection problem has been written as a model selection problem, the aim of this section is to propose a criterion to select the model M_{τ_m} that contains the true additive outliers in the series. Note that the candidate models include the one without outliers, M_{τ_0}, the T models with one outlier, M_{τ_1}, and so on. In total, there are $\binom{T}{m}$ candidate models with m outliers covering all the possible outlier locations. Thus, assuming that the number of outliers has an upper bound, $m_{\max} < T$, the total number of candidate models is given by

$$\binom{T}{0} + \binom{T}{1} + \cdots + \binom{T}{m_{\max}}. \tag{13.3.1}$$

This section proposes a model selection criterion for selecting the model M_{τ_m} by a modification of the BIC that includes an additional term that may be useful for outlier detection.

The BIC is derived after approximating the posterior distributions of the candidate models, denoted as $p(M_{\tau_m}|y)$. This is given by (see Claeskens and Hjort (2008)),

$$p(M_{\tau_m}|y) = \frac{p(M_{\tau_m})\mathcal{L}(M_{\tau_m}|y)}{f(y)}, \tag{13.3.2}$$

where $p(M_{\tau_m})$ is the prior probability of model M_{τ_m}, $\mathcal{L}(M_{\tau_m}|y)$ is the marginal likelihood for model M_{τ_m} given by

$$\mathcal{L}(M_{\tau_m}|y) = \int \mathcal{L}\left(\rho_{\tau_m}|y, M_{\tau_m}\right) p\left(\rho_{\tau_m}|M_{\tau_m}\right) d\rho_{\tau_m}, \tag{13.3.3}$$

with $p(\rho_{\tau_m}|M_{\tau_m})$, the prior probability of the parameters given the model M_{τ_m}, and $f(y)$ is the unconditional likelihood of y given by

$$f(y) = \sum_{j=0}^{m_{\max}} \sum_{\tau_j} p(M_{\tau_j})\mathcal{L}(M_{\tau_j}|y).$$

From a Bayesian point of view, and taking into account that $f(y)$ is constant for all the models, in order to compute $p(M_{\tau_m}|y)$, it is required to give prior probabilities to the models M_{τ_m} and to compute the marginal likelihood for each model M_{τ_m}, $\mathcal{L}(M_{\tau_m}|y)$. Therefore, calculation of the posterior probabilities in (13.3.2) requires specification of the priors of the models and parameters, and integration over the parameter space. However, obtaining an analytical expression for $\mathcal{L}(M_{\tau_m}|y)$ is infeasible. Alternatively, a second-order expansion of the log-likelihood function, $\ell_{\tau_m}(\rho_{\tau_m}) = \log \mathcal{L}(\rho_{\tau_m}|y, M_{\tau_m})$, around the MLEs, $\widehat{\rho}_{\tau_m}$, leads to the following Laplace approximation to the integral in (13.3.3); see Claeskens and Hjort (2008),

$$\mathcal{L}(M_{\tau_m}|y) = \left(\frac{2\pi}{T}\right)^{\frac{p_m}{2}} \exp\left(\ell_{\tau_m}(\widehat{\rho}_{\tau_m})\right) p\left(\widehat{\rho}_{\tau_m}|M_{\tau_m}\right) \left|H_{\tau_m}\left(\widehat{\rho}_{\tau_m}\right)\right|^{-\frac{1}{2}} + O\left(T^{-\frac{p_m}{2}-1}\right),$$

where $p\left(\widehat{\rho}_{\tau_m}|M_{\tau_m}\right)$ is the prior of the parameters given the model, and $\ell_{\tau_m}(\widehat{\rho}_{\tau_m})$ and $H_{\tau_m}(\widehat{\rho}_{\tau_m})$ are, respectively, the log-likelihood and the Hessian matrix of $T^{-1}\ell_{\tau_m}(\rho_{\tau_m})$, all evaluated at $\widehat{\rho}_{\tau_m}$. Therefore, (13.3.2) can be written as follows:

$$p(M_{\tau_m}|y) = \frac{p(M_{\tau_m})}{f(y)}\left[\left(\frac{2\pi}{T}\right)^{\frac{p_m}{2}} \exp\left(\ell_{\tau_m}(\widehat{\rho}_{\tau_m})\right) p\left(\widehat{\rho}_{\tau_m}|M_{\tau_m}\right) \left|H_{\tau_m}\left(\widehat{\rho}_{\tau_m}\right)\right|^{-\frac{1}{2}} + O\left(T^{-\frac{p_m}{2}-1}\right)\right], \tag{13.3.4}$$

which also depends on the prior probabilities of the models, $p(M_{\tau_m})$, and on the unconditional likelihood, $f(y)$. Taking logarithms, (13.3.4) leads to,

$$\log p(M_{\tau_m}|y) = \ell_{\tau_m}(\widehat{\rho}_{\tau_m}) + \frac{p_m}{2}\log\frac{2\pi}{T} - \frac{1}{2}\log\left|H_{\tau_m}(\widehat{\rho}_{\tau_m})\right| + \log p(\widehat{\rho}_{\tau_m}|M_{\tau_m}) + \log p(M_{\tau_m}) - \log f(y) + O(T^{-1}). \tag{13.3.5}$$

Following Claeskens and Hjort (2008), the dominant terms in (13.3.5) are the first two, which are of sizes $O_P(T)$ and $\log T$, respectively, while the others are $O_P(1)$. The usual BIC approximation of the posterior probability in (13.3.5) is based on assuming uniform prior probabilities for all the candidate models. Thus, the prior probability of model M_{τ_m} under the BIC approximation is given by

$$p_{\text{BIC}}(M_{\tau_m}) = \frac{1}{\binom{T}{0} + \binom{T}{1} + \cdots + \binom{T}{m_{\max}}},$$

which is independent of the number of outliers, m. Now, taking uniform prior probabilities for the parameters of the models and ignoring all the lower order terms, minus two times (13.3.5) leads to the BIC, which selects the model M_{τ_m} that minimizes

$$\text{BIC}(M_{\tau_m}) = -2\ell_{\tau_m}\left(\widehat{\rho}_{\tau_m}\right) + p_m \log T. \tag{13.3.6}$$

However, note that the prior probability of the number of outliers used by the BIC approximation is given by

$$p_{\text{BIC}}(m) = \sum_{\tau_m,\ m \text{ fixed}} p_{\text{BIC}}(M_{\tau_m}) = \frac{\binom{T}{m}}{\binom{T}{0} + \binom{T}{1} + \cdots + \binom{T}{m_{\max}}}. \tag{13.3.7}$$

As a consequence, when $m_{\max} \ll T/2$, i.e., when the maximum possible number of outliers is small compared with $T/2$, as is expected in most real time series applications, the model with the largest prior probability is the model with the largest possible number of outliers. Indeed, the prior probabilities assigned by the BIC in (13.3.7) are an increasing function of the number of outliers m, which may be unreasonable. For instance, note that $p_{\text{BIC}}(1)$ and $p_{\text{BIC}}(2)$ are T and $T(T-1)/2$ times larger than $p_{\text{BIC}}(0)$.

Next, an alternative criterion to the BIC in (13.3.6) is proposed. This is called *BICUP* (for BIC with uniform prior), and it is based on penalizing for the possible number of outliers. This leads to a uniform prior distribution over the number of outliers. Then, taking equal prior probabilities for all the models with the same number of outliers, the prior probability of the model M_{τ_m} under the *BICUP* approximation is given by

$$p_{BICUP}(M_{\tau_m}) = \frac{1}{1+m_{\max}} \frac{1}{\binom{T}{m}}.$$

Now, taking uniform prior probabilities for the parameters of the models, and after deleting constants and low order terms, minus two times (13.3.5) leads to the *BICUP* for outlier detection, which selects the model M_{τ_m} that minimizes

$$BICUP(M_{\tau_m}) = -2\ell_{\tau_m}(\widehat{\rho}_{\tau_m}) + p_m \log T + 2\log \binom{T}{m}. \tag{13.3.8}$$

Note that the expression of the *BICUP* in (13.3.8) is similar to the expression of the BIC in (13.3.6) except for the last term, which shows an additional penalization for models that increases with m. Consequently, the *BICUP* naturally incorporates the information about the number of models for different numbers of outliers. The prior probability of the number of outliers taken by the *BICUP* approximation is given by

$$p_{BICUP}(m) = \sum_{\tau_m,\ m \text{ fixed}} p_{BICUP}(M_{\tau_m}) = \frac{1}{1+m_{\max}},$$

which has the attractive property that the probability of having an additional additive outlier does not depend on the global number of outliers, since the prior ratio,

$$\frac{p_{BICUP}(M_{\tau_{m+1}})}{p_{BICUP}(M_{\tau_m})} = 1,$$

is independent of m. Then, all the possible numbers of outliers are equally probable *a priori*. Figure 13.1 shows the last term in (13.3.8) as a function of m and T. In particular, the additional term increases with m and/or T, so that the penalization is larger than the BIC penalization for large values of m and T.

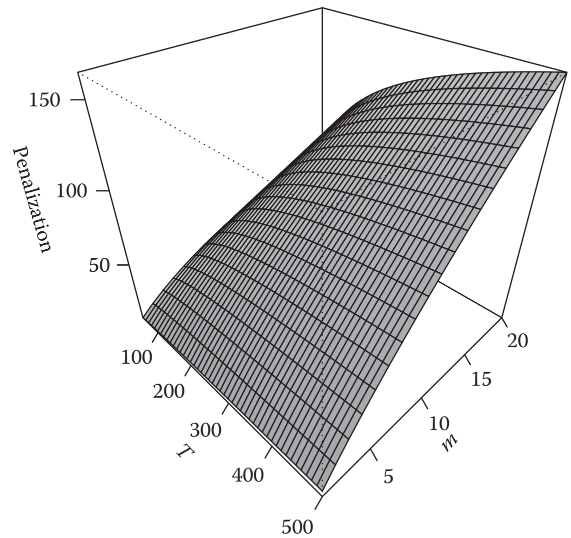

FIGURE 13.1
Last term of *BICUP* as a function of T and m.

In summary, the model M_{τ_m} is selected as the one that provides the minimum value of the proposed *BICUP*. Note that the model selection determines the number of outliers, m, their locations, τ_m, and the MLEs, $\widehat{\rho}_{\tau_m}$, of the model parameters. Additionally, the model selection criterion for outlier detection in (13.3.8) provides approximations of the posterior probabilities $p(M_{\tau_m}|y)$. More precisely, the approximated value of $p(M_{\tau_m}|y)$ is given by

$$p_{BICUP}(M_{\tau_m}|y) = \frac{\exp\left(-\dfrac{BICUP(M_{\tau_m})}{2}\right)}{\sum_{j=0}^{m_{\max}} \sum_{\tau_j} \exp\left(-\dfrac{BICUP\left(M_{\tau_j}\right)}{2}\right)}.$$

Also, pairwise comparison of two models can be done using the *BICUP* approximation of the posterior odds for model M_{τ_m} against M_{τ_n}, which is given by

$$\begin{aligned} o_{BICUP}\left(M_{\tau_m}, M_{\tau_n}|y\right) &= \frac{p_{BICUP}(M_{\tau_m}|y)}{p_{BICUP}\left(M_{\tau_n}|y\right)} \\ &= \exp\left(\frac{BICUP(M_{\tau_n}) - BICUP(M_{\tau_m})}{2}\right), \end{aligned} \qquad (13.3.9)$$

which only requires computation of the values of the *BICUP* for models M_{τ_m} and M_{τ_n}.

13.4 Procedure for Detecting Potential Outliers

There is an additional problem in computing the value of the *BICUP*. As noted in Section 13.2, the number of candidate models, given in (13.3.1), may be huge even for small values of T and $m_{\max}$. Consequently, getting the values of the proposed criterion for all the possible candidate models is a computationally expensive problem. This section proposes a procedure for reducing the number of models for which the criterion should be computed. The procedure is based on that proposed by Peña and Tiao (1992) for defining Bayesian robustness functions in linear models. The idea is to split the time series observations into two groups: the first would include observations that have high potential of being outliers, while the second includes the observations that should be discarded as outliers beyond any reasonable doubt. If T_1 is the number of observations in the first group, then, instead of computing the value of the proposed criterion for all the candidate models, it is possible to compute it for all the models, which include as outliers all the combinations of the T_1 observations in the first group. Thus, the number of candidate models reduces to,

$$\binom{T_1}{0} + \binom{T_1}{1} + \cdots + \binom{T_1}{T_1},$$

which is a much smaller number than the one in (13.3.1).

Obviously, the key point in the procedure is to split the time series observations into these two groups. Due to the masking and swamping effects, the groups cannot be made by simply computing the value of the proposed criterion for the T models M_{τ_1}. Alternatively, the following approach is considered. Let A_r be the event "the observation y_r is an outlier given the y." Then, the probability of two observations, y_r and y_s, being outliers can be written as follows:

$$P(A_r \cap A_s) = P(A_r|A_s)P(A_s). \tag{13.4.1}$$

If y_r and y_s are nonoutlying time points, the probability in (13.4.1) is approximately given by $P(A_r)P(A_s)$ because $P(A_r|A_s) \simeq P(A_r)$. However, if y_s is an outlier, $P(A_r|A_s)$ will be very different than $P(A_r)$ because the correct detection of an outlier will affect the value of $P(A_r)$, and $P(A_r \cap A_s)$ will be quite different from $P(A_r)P(A_s)$. As a consequence, a way to distinguish potential outliers is to examine the values of the differences, $P(A_r \cap A_s) - P(A_r)P(A_s)$.

However, computation of these probabilities is also a difficult task because they involve a large number of probabilities. As an alternative, it is possible to use the approximated posterior odds given in (13.3.9). The idea is to build the interactions matrix with elements,

$$\begin{aligned} &d_{BICUP}(r,s) \\ &\quad = |o_{BICUP}(M^{r,s}_{\tau_2}, M_{\tau_0}|y) - o_{BICUP}(M^{r}_{\tau_1}, M_{\tau_0}|y)o_{BICUP}(M^{s}_{\tau_1}, M_{\tau_0}|y)|, \end{aligned}$$

for $r, s = 1, \ldots, T$, where $o_{BICUP}(M_{\tau_2}^{r,s}, M_{\tau_0}|y)$ is the *BICUP* approximation of the posterior odds of the model that assumes that y_r and y_s are outliers against M_{τ_0}, and $o_{BICUP}(M_{\tau_1}^{r}, M_{\tau_0}|y)$ and $o_{BICUP}(M_{\tau_1}^{s}, M_{\tau_0}|y)$ are the *BICUP* approximations of the posterior odds of the model that assumes that, on the one hand, y_r and, on the other hand, y_s, are outliers, against M_{τ_0}. If y_r is an outlier, the values $d_{BICUP}(r, \cdot)$ are expected to have relatively large values. Indeed, if there are other outliers masked by y_r, these will show up as large values in the distribution of $d_{BICUP}(r, \cdot)$. Thus, large values of $d_{BICUP}(r, \cdot)$ will indicate outliers, and relatively large values in a column, possible masking between these points. Therefore, a procedure for pointing out potential outliers is the following:

1. Compute the values of $o_{BICUP}(M_{\tau_1}^{r}, M_{\tau_0}|y)$ for all the models with an outlier at time point $r = 1, \ldots, T$. Let $m(o_{BICUP})$ and $sd(o_{BICUP})$ be the mean and the standard deviation of the values of $o_{BICUP}(M_{\tau_1}^{r}, M_{\tau_0}|y)$. Then, include in the set of potential outliers those points that satisfy
$$o_{BICUP}\left(M_{\tau_1}^{r}, M_{\tau_0}|y\right) \geq m(o_{BICUP}) + 3 \times sd(o_{BICUP}). \quad (13.4.2)$$

2. Compute the values of $d_{BICUP}(r, s)$ for all the models with two outliers at the time points $r, s = 1, \ldots, T$. Let $m(d_{BICUP})$ and $sd(d_{BICUP})$ be the mean and the standard deviation of the values of $d_{BICUP}(r, \cdot)$. Then, include in the set of potential outliers those points y_s that satisfy
$$d_{BICUP}(r, s) \geq m(d_{BICUP}) + 5 \times sd(d_{BICUP}). \quad (13.4.3)$$

Several comments on this procedure are in order. First, once the procedure provides the set of potential outliers, the values of the *BICUP* for all the models included in this set are computed and the model that gives the minimum value of the *BICUP* is the selected model. Then, estimation of the model parameters and outlier sizes is made jointly through the MLEs. Second, note that the use of the procedure avoids the problem of choosing the value of $m_{\max}$, i.e., the maximum number of outliers allowed, because it is only required to compute the value of the criterion for those models that include potential outliers. Therefore, the number of potential outliers can be seen as the value of $m_{\max}$. Third, the values 3 and 5 have been chosen following the suggestions in Peña and Tiao (1992). Indeed, in our experience with simulated time series, these values provide a number of potential outliers equal to or slightly larger than the true number of outliers. Therefore, these values contribute in an appropriate way to establishing the observations that are suspected to be outliers and to give an accurate estimate of the maximum number of possible outliers m_{max}. This is important because taking smaller values may lead to a large group of potential outliers that includes false outliers, which may lead to estimation problems, while taking larger values may lead to a small group of potential outliers not including all the true outliers. In those unexpected situations in which the number of candidate outliers given by the procedure is large, we can use the median and MAD (median of the absolute deviations from the sample median) instead of the mean and standard deviation. In this situation, following Peña (2005), we can consider as heterogeneous observations those that deviate from the median by more than 4.5 times the MAD. Fourth, it may appear that the *BICUP* interactions matrix is only able to point out the presence of two outliers and that higher order interactions should be analyzed. However, our simulation experiments showed us that this not the case. An example of this can be seen in Section 13.5.

13.5 Examples

This section illustrates the performance of the proposed outlier detection methodology for a simulated series and for the time series of logarithms of the monthly total retail sales in the United States.

13.5.1 Outlier detection for a simulated time series

A simulated time series contaminated with three additive outliers is built as follows. First, a series with sample size $T = 100$ is simulated from the seasonal ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ model given by

$$(1 - B^{12})(1 - B)x_t = (1 + 0.5B^{12})(1 + 0.4B)\epsilon_t,$$

where ϵ_t follows a Gaussian distribution with zero mean and standard deviation $\sigma = 0.7071$. Then, the series $(x_1, \ldots, x_T)$ is contaminated with three additive outliers at time points $t_1 = 50$, $t_2 = 51$, and $t_3 = 52$ and outlier sizes $w_{50} = 3$, $w_{51} = -3$, and $w_{52} = 3$, respectively. Thus, the outlier magnitudes are around 4.25 times the error standard deviation. The contaminated time series is, then, given by

$$y_t = x_t + 3I_t^{(50)} - 3I_t^{(51)} + 3I_t^{(52)},$$

for $t = 1, \ldots, T$. Both the outlier-free and the contaminated time series are shown in Figure 13.2. Note that the outlier effects are almost imperceptible in the plot. However, the additive outliers produce large effects in parameter estimation. This is illustrated in Table 13.1, which includes the estimates of the parameters of two seasonal ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ models fitted to the contaminated series y_t. The table shows that the parameter estimates ignoring the presence of the additive outliers are very different than the true model parameters, whereas when the model includes the three outliers the estimation is accurate.

Next, we apply the procedure described in Section 13.4 for selecting potential outliers to the contaminated series $(y_1, \ldots, y_T)$. The first step of the

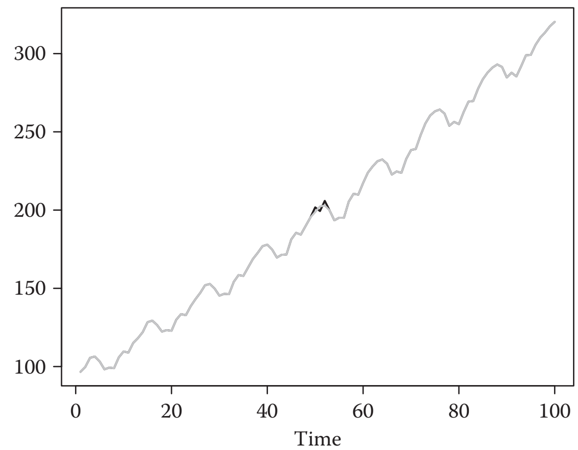

FIGURE 13.2
Simulated time series with (in black) and without (in gray) outliers.

procedure consists of computing the values of the *BICUP* approximated posterior odds for models with one outlier, i.e., $M^t_{\tau_1}$ for $t = 1, \ldots, T$, against M_{τ_0}. Figure 13.3 shows these values. The straight horizontal line in the plot is the height of the threshold in (13.4.2). The plot in Figure 13.3 shows that the observation at time point $t = 51$ is labeled as a potential outlier because the value of its *BICUP* approximated posterior odds is much larger than the threshold. Importantly, note that the observations at time points $t = 50$ and

TABLE 13.1
Estimated parameters and standard errors in the simulated series contaminated by three outliers.

True parameter values	Estimated parameter values: Model without the outliers	Estimated parameter values: Model with the outliers
$\theta_1 = -0.4$	0.378	−0.456
	(0.092)	(0.125)
$\Theta_1 = -0.5$	0.220	−0.408
	(0.135)	(0.153)
$w_{50} = 3$	—	2.814
		(0.58)
$w_{51} = -3$	—	−2.973
		(0.785)
$w_{52} = 3$	—	2.84
		(0.568)
$\sigma = 0.707$	1.373	0.722

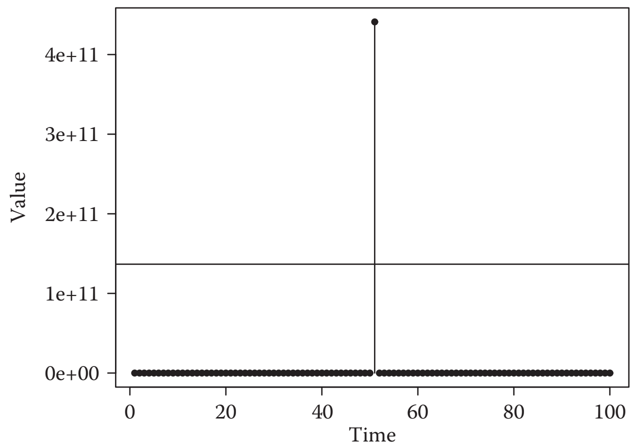

FIGURE 13.3
BICUP approximated posterior odds for models with one outlier in the simulated series contaminated with three outliers.

$t = 52$ are not labeled as potential outliers in this first step, maybe because they are masked by the outlier at $t = 51$. This example shows the need for the second step in the algorithm described in Section 13.4.

The second step of the procedure starts by computing the *BICUP* interactions matrix and the corresponding thresholds for the 100 rows of the matrix given in (13.4.3). The three observations at $t = 50$, $t = 51$, and $t = 52$ are labeled as potential outliers by the procedure. More precisely, on the one hand, the rows 1 to 49 and 53 to 100 of the interactions matrix point out that the observation at $t = 51$ is a potential outlier. On the other hand, the rows 50 and 51, and then 51 and 52, point out that the observations at $t = 52$ and $t = 50$ are also potential outliers, respectively. Figure 13.4 shows the row numbers 50, 51, and 52 of the *BICUP* interactions matrix. The straight horizontal lines in the plots are the height of the corresponding thresholds given in (13.4.3). Note that there are no nonoutlier observations pointed out as potential outliers.

The last step of the proposed methodology is to compute the values of the *BICUP* for only those models that incorporate the potential outliers at time points $t = 50$, $t = 51$, and $t = 52$. Table 13.2 shows these values for all the models considered. As can be seen, the *BICUP* selects the model with the true outliers. Once the outliers have been detected, inference on the model parameters can be performed through the MLEs. The second column in Table 13.1 shows the parameter estimates of the model that incorporates the true outliers. Note that the parameter estimates of the seasonal $\text{ARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ fitted to the contaminated series y_t and the outlier size estimates are very close to their real values.

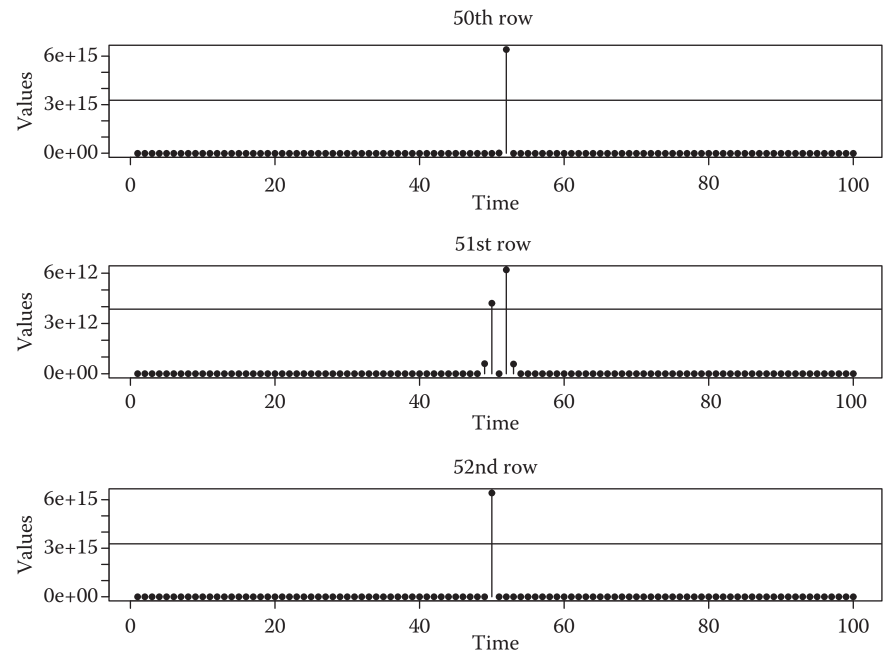

FIGURE 13.4
Rows of the *BICUP* interactions matrix in the simulated series contaminated with three outliers.

Finally, although the proposed procedure is only suited for series with additive outliers, we compare the previous results with the results obtained with the seasonal adjustment software X-12-ARIMA developed by the U.S. Census Bureau. The X-12-ARIMA software, besides fitting the seasonal ARIMA model, searches for three outlier types—additive outliers, level changes and transitory changes—using a sequential detection procedure. In this case, the X-12-ARIMA software detects a level shift at $t = 50$ with estimated size 3.0391, an additive outlier at $t = 51$ with estimated size -5.8669, and a level shift at $t = 53$ with estimated size -2.8975. Note that the estimated size of the second level shift is very close to the estimated size of the first level shift but with negative sign. Therefore, the second level shift somehow cancels the effect of the first level shift.

TABLE 13.2
Values of the criteria for candidate models in the simulated series contaminated with three outliers.

τ_m	$(-)$	(50)	(51)	(52)	$(50, 51)$	$(50, 52)$	$(51, 52)$	$(50, 51, 52)$
BICUP	321.3	316.6	267.6	315.9	268.5	248.5	269	**246.8**

13.5.2 Outlier detection for the logarithms of the monthly total retail sales in the United States

The performance of the proposed methodology is illustrated by analyzing the logarithms of the monthly total retail sales in the United States. The time series, which starts in January 1992 and ends in December 2007, so that it consists of $T = 192$ data points, is plotted in Figure 13.5. The series has clearly seasonal behavior. In order to account for trading-day effects, we include seven regressor variables in the model. The first six variables are defined as $r_{1t} = (\textit{no. of Mondays}) - (\textit{no. of Sundays})$ in month $t, \ldots, r_{6t} = (\textit{no. of Saturdays}) - (\textit{no. of Sundays})$ in month t, along with a variable defined as $r_{7t} = \textit{length of month } t$. Then, a seasonal ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ model plus the seven regressors is fitted to the time series. The autocorrelation function of the residual series does not show serial dependence, so that the fit appears to be appropriate. Table 13.3 includes the estimates of the parameters of two seasonal ARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ models with trading-day effects fitted to the time series. In particular, the second column in Table 13.3 shows the MLEs of the parameters of this model along with their standard errors.

Next, we apply the procedure described in Section 13.4 for selecting potential outliers in the logarithm of the monthly total retail sales series. The first step of the procedure consists of computing the values of the *BICUP* approximated posterior odds for models with one outlier. Figure 13.6 shows these values. The straight horizontal line in the plot is the height of the threshold in (13.4.2). The plot in Figure 13.6 shows that the observation at October 2001 is labeled as a potential outlier because the value of its *BICUP* approximated

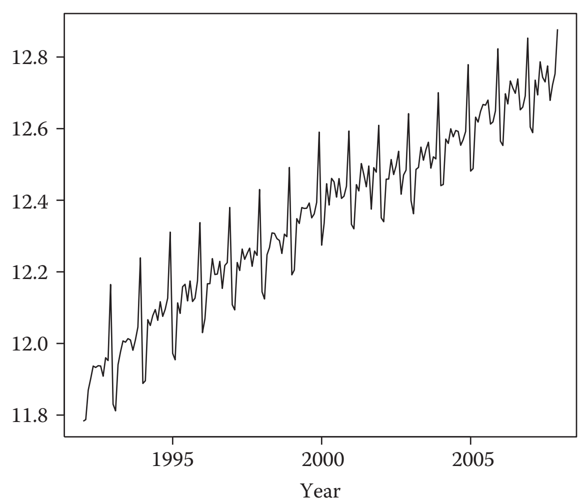

FIGURE 13.5
Logarithm of the monthly total retail sales in the United States.

TABLE 13.3
Estimated parameters and standard errors in the monthly total retail sales series.

	Estimated parameter values	
Parameter	**Seasonal ARIMA**	**Outliers at October 2001 and May 2005**
θ_1	0.570	0.426
	(6.73×10^{-2})	(9.38×10^{-2})
Θ_1	0.602	0.547
	(5.83×10^{-2})	(5.84×10^{-2})
β_1	0.5×10^{-3}	-0.6×10^{-3}
	(1.8×10^{-3})	(1.5×10^{-3})
β_2	1.9×10^{-3}	1.8×10^{-3}
	(1.8×10^{-3})	(1.4×10^{-3})
β_3	1.1×10^{-3}	0.5×10^{-3}
	(1.9×10^{-3})	(1.5×10^{-3})
β_4	5.7×10^{-3}	6.7×10^{-3}
	(1.8×10^{-3})	(1.5×10^{-3})
β_5	6.7×10^{-3}	6.6×10^{-3}
	(1.8×10^{-3})	(1.4×10^{-3})
β_6	-1.2×10^{-3}	-1.2×10^{-3}
	(1.8×10^{-3})	(1.5×10^{-3})
β_7	4.11×10^{-2}	4.02×10^{-2}
	(5.8×10^{-3})	(0.47×10^{-3})
w_{118}	—	5.38×10^{-2}
		(8.8×10^{-3})
w_{161}	—	-4.4×10^{-2}
		(8.7×10^{-3})
σ	1.33×10^{-2}	1.15×10^{-2}

posterior odds is larger than the threshold. None of the values of the *BICUP* approximated posterior odds for the rest of observations is close to the corresponding threshold.

Then, the second step of the procedure starts by computing the *BICUP* interactions matrix. Figure 13.7 shows the row numbers 118 and 161 of this matrix. The straight horizontal lines in the plots are the height of the corresponding threshold given in (13.4.3). These plots show that the observations in October 2001 and May 2005 are labeled as potential outliers. Indeed, these two observations are labeled as potential outliers in many of the rows of the interaction matrix that are not shown here. No more observations are labeled as potential outliers.

The final step of the proposed methodology is to compute the values of the *BICUP* for only those models that incorporate the potential outliers at October 2001 and May 2005. Table 13.4 shows these values for all the models considered. The *BICUP* selects the model with outliers in October 2001 and

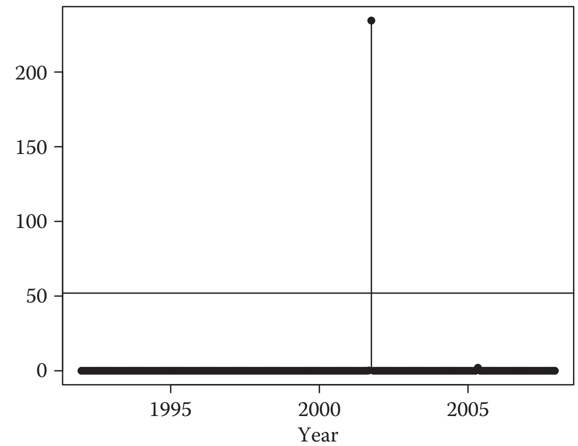

FIGURE 13.6
BICUP approximated posterior odds for models with one outlier in the logarithm of the monthly total retail sales series in the United States.

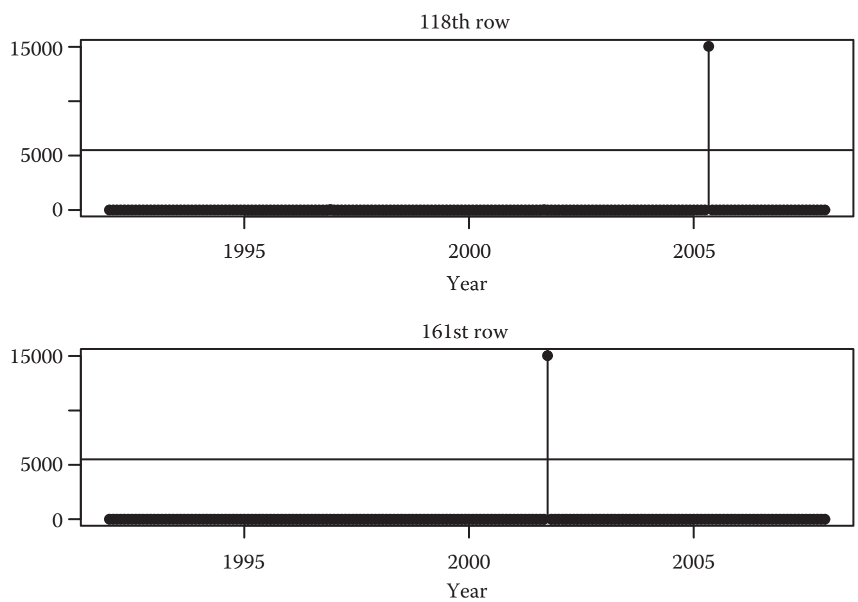

FIGURE 13.7
118th and 161st rows of *BICUP* interactions matrix in the logarithm of the monthly total retail sales series in the United States.

TABLE 13.4
Values of the criteria for candidate models in the logarithm of the monthly total retail sales series.

τ_m	(−)	(118)	(161)	(118, 161)
BICUP	−1015.2	−1026.2	−1016.7	−**1034.5**

May 2005. Thus, the proposed criterion provides a model for the logarithm of the monthly total retail sales in the United States that incorporates two outliers. The third column in Table 13.3 shows the MLEs of the parameters of this model. This includes the outlier size estimates of the two outliers, $\widehat{w}_{118}$ and $\widehat{w}_{161}$. Apparently, the first/second outlier produced an increase/decrease in the monthly total retail sales series.

Finally, as in the simulated example, we compare the previous results with the results produced by X-12-ARIMA. For that, we include regressor variables to model trading-day and Easter effects. For this series, the X-12-ARIMA software only detects an additive outlier in October 2001 with estimated size 0.0527, which is very close to the estimated value in our final model with two outliers. X-12-ARIMA does not detect the observation in May 2005 as an outlier.

Acknowledgments

The authors would like to thank the excellent comments made by the editors and two referees, which have led to a much improved chapter. The first author acknowledges the financial support by Ministerio de Ciencia e Innovación project MTM2008-03010. The second author acknowledges the financial support by Ministerio de Educación y Ciencia project SEJ2007/64500/ECON4438.

References

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, 21(1):243–247.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, 267–281. Akademiai Kiadó, Budapest.

Bianco, A. M., Garcia Ben, M., Martínez, E. J., and Yohai, V. J. (2001). Outlier detection in regression models with ARIMA errors using robust estimates. *Journal of Forecasting*, 20(8):565–579.

Chang, I., Tiao, G. C., and Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204.

Chen, C. and Liu, L. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421):284–297.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. New York: North-Holland.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society Series B*, 34(3):350–363.

Gómez, V. and Maravall, A. (1994). Estimation, prediction, and interpolation for nonstationary series with the Kalman filter. *Journal of the American Statistical Association*, 89(426):611–624.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society Series B*, 41(2):190–195.

Harvey, A. C. and Phillips, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, 66(1):49–58.

Hurvich, C. M. and Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307.

Justel, A., Peña, D., and Tsay, R. S. (2001). Detection of outlier patches in autoregressive time series. *Statistica Sinica*, 11(3):651–673.

Kohn, R. and Ansley, C. F. (1985). Efficient estimation and prediction in time series regression models. *Biometrika*, 72(3):694–697.

Luceño, A. (1998). Detecting possibly non-consecutive outliers in industrial time series. *Journal of the Royal Statistical Society Series B*, 60(2):295–310.

McCulloch, R. E. and Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis*, 15(2):235–250.

Peña, D. (2005). A new statistic for influence in linear regression. *Technometrics*, 47(1):1–12.

Peña, D. and Tiao, G. C. (1992). Bayesian robustness functions for linear models. In *Bayesian Statistics 4*, 365–388. Oxford University Press.

Perron, P. and Rodriguez, G. (2003). Searching for additive outliers in nonstationary time series. *Journal of Time Series Analysis*, 24(2):193–220.

Sánchez, M. J. and Peña, D. (2003). The identification of multiple outliers in ARIMA models. *Communications in Statistics, Theory and Methods*, 32(6):1265–1287.

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

Tsay, R. S. (1986). Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, 81(393):132–141.

Tsay, R. S. (1988). Outliers, level shifts and variance changes in time series. *Journal of Forecasting*, 7(1):1–20.