ROBUST TIME SERIES ESTIMATION VIA WEIGHTED LIKELIHOOD.
SOME PRELIMINARY RESULTS

C. Agostinelli

2001.3

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Marzo 2001

Robust time series estimation via weighted likelihood

Some preliminary results

Claudio Agostinelli *
Department of Statistics
University of Padova
35121 Padova, Italia

February 21, 2001

Abstract

In this paper we discuss a preliminary results on the construction of a weighted likelihood procedure for robust estimation of the unknown parameters of an autoregressive-moving average model. Two types of outliers, i.e., additive and innovations be take into account without knowing their number, position or intensity. A classification procedure based on a selection criterion is used to identified the most useful pattern rappresentation of the outliers to gain efficiency and robustness. Two examples are reported.

Key words: Additive outliers, Autoregressive-moving average model, Innovations outlier, Outliers classification procedure, Robust estimation, Weighted likelihood.

*Address of the author: Claudio Agostinelli, Dipartimento di Statistica, Via San Francesco n. 33, Università di Padova, 35121 Padova, Italia. e-mail: claudio@stat.unipd.it

1 Introduction

In this paper we will consider the problem of robust estimation of the parameters of a general stationary invertible $ARMA(p,q)$ model

$$\Phi(B)Z_t = \theta_0 + \Theta(B)a_t \tag{1}$$

where $\{a_t\}$ are distributed as a white noise processes (often with density $\mathcal{N}(0,\sigma_a^2)$, with nuisance scale parameter σ_a^2), $\Phi(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$, $\Theta(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$, θ_0 is the mean of the process and B is the backward shift operator such that $Z_{t-1} = BZ_t$. Further, let $\beta = \{\Phi(B), \Theta(B), \theta_0\}$ the vector of unknown parameters.

Outliers data are observations that departure from the model of the bulk of the data. This kind of observations can lead to a significant impact on the effectiveness of the standard methodology for time series with respect to model selection, estimation and forecasting. Maddala and Yin (1997) presents an update review of this subject in the literature. Two main fields are present: i) outlier detection and accomodation and ii) robust methods. In the first field we find a growing presence of papers: Fox (1972) first introduced a classification methods and suggested likelihood ratio tests for outliers when their location and type are known. Several extensions are made to this work, among others, Tsay (1986, 1988) propose an iterative procedure to identify outliers (a recent review of Tsay's work together with a comparison of other methods is presented in Vaage (2000)) and Chen and Liu (1993) extend Tsay's works to the presence of multiple outliers. Both methods are based on least squares and/or maximum likelihood, and hence they will not be very stable under the presence of outliers. Franses and Lucas (1995) use robust estimation methods to avoid this problem. Robust methods for time series were difficult to developed since the class of M-estimator are not robust in this case and the masking issues due to the presence of multiple outliers. Bustos and Yohai (1986) developed estimators based on residual autocovariance and on truncated residual autocovariances, Masarotto (1987) extended the notion of maximum likelihood of order m (Azzalini, 1983) to M-estimator in order to bound the influence of outliers observations.

We will introduce an algorithm based on weighted likelihood in order to obtain a robust estimation procedure of the unknown parameters β and σ_a^2.

In Section 2 we review the outliers type considered in literature, in Section 3 we introduce the weighted likelihood estimating equations for Arma models and the weighted likelihood estimating equations of order m; Section 4 presents the algorithm procedure while Section 5 presents two examples. Finally in the Section 6 we discuss the properties of the introduced method.

2 Outliers type in time series

In literature (see among others Barnett and Tobis, (1994)) are often define two different type of contamination: additive outliers (AO) and innovations outliers (IO). Some other type such as level change (LC) and transient change (TC) are also considered as in Tsay (1988), but we will not deal with them here since these kinds of contaminations are structural and possible permanent change in the data generation model and so they can not be assimilate to outliers observations. In these cases a mixture model could be useful. Any contaminated time series Y_t can be represented as a function of the uncontaminated time series Z_t (following some arma model) and a component taking into account the outliers behavior as

$$Y_t = \sum_{j=1}^{n} h_j v_j(B) \xi_t^{(t_j^*)} + Z_t \tag{2}$$

where

- $v_j(B) = 1$ for an AO,
- $v_j(B) = [\Theta(B)/\Phi(B)]$ for an IO,

at time $t = t_j^*$, h_j represents the magnitude of the outliers, $\xi_t^{(t^*)} = 1$ if $t = t^*$, and $= 0$ otherwise and n is the number of outliers.

Note that while an additive outlier will affect only the corresponding observation, an innovations outlier will be propagated according to $\Theta(B)/\Phi(B)$.

Since we can only have realizations from Y_t when we apply the arma model of Z_t to the observed time series we have the following observed residuals r_t as function of the error a_t

$$r_t = \frac{\Phi(B)}{\Theta(B)} Y_t = a_t + \sum_{j=1}^{n} h_j \tilde{v}_j(B) \xi_t^{(t_j^*)} \tag{3}$$

where

- $\tilde{v}_j(B) = \Phi(B)/\Theta(B)$ for an AO,
- $\tilde{v}_j(B) = 1$ for an IO,

Hence, while an IO affect only one r_t, an AO affect all observations after r_t according to $\tilde{v}_j(B)$ too.

Further, it is important to note that every outliers sequence can not be identified uniquely, since it is always possible to find a different sequence of outliers (maybe with infinite elements) that generate the same observed time series. To illustrate, let consider an AR(1) process $Z_t = \phi_1 Z_{t-1} + a_t$ and a realization from it z_t. When an additive outlier occur at time t^* with magnitude $h_{t^*}^{AO}$ we observe the time series

$$y_t = z_t + h_{t^*}^{AO} \ \xi_t^{(t^*)}$$

but the same time series y_t could be observe by the presence of two innovations outliers in t^* and t^*+1 such that $h_{t^*}^{IO} = h_{t^*}^{AO}$ and $h_{t^*+1}^{IO} = -h_{t^*}^{AO}\phi_1$. The same result can be easily generalized for every process and for every outliers pattern. Hence build a classification procedure that correctly identified the outliers generation process can not be construct. On the other hand when a time series contains an outlier that could be "view" as an additive outlier, the number of spurious observations could be infinity in the presence of a moving average part in the model and the weighted likelihood estimator would certainly breakdown. This behavior is well known in classical M-estimator too. Hence our goal is to look for a classification procedure such that an observation, say y_{t^*} is classify as an additive outliers only if it cause outlying behavior on some of the successive residuals $r_{t^*+1}, \cdots, r_T$. To bound its effects we will replace it by its best linear predictor. Replacing an outliers observations, say y_{t^*}, with its best linear predictor when y_{t^*} does not affect any of the future residuals $r_{t^*+1}, \cdots, r_T$ means to produce a new artificial outlier and reduce the efficiency and the robust properties of the estimation method.

3 Weighted Likelihood Estimating Equations for Arma models

The definition of Weighted Likelihood Estimating Equations (WLEE) was first derive by Lindsay (1994) for discrete distributions. Markatou, Basu and Lindsay (1998) extend the methods to the continuous models. Agostinelli (1998) define the WLEE for the regression models, he also developed robust model selection criteria (see the literature in Agostinelli (2000)) while Agostinelli and Markatou (2000) define robust tests analogous to the classical test based on likelihood functions. Wide applicability, first order efficiency and positive breakdown points are the main features of this approach.

Several computational methods are available to perform the estimation of unknown parameters in the classical likelihood setting: conditional likelihood, unconditional likelihood, exact maximum likelihood. To illustrate we define the conditional weighted likelihood estimating equations, the others are obtain in similar way.

The joint probability density of $a = \{a_1, a_2, \cdots, a_T\}$ is given by

$$P(a|\Phi, \theta_0, \Theta, \sigma_a^2) = \left(2\pi\sigma_a^2\right)^{(-n/2)} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{T} a_t^2\right]$$

Rewriting 1 as

$$a_t = -\theta_0 + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + Z - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p},$$

we can write down the likelihood function of the parameters $\{\Phi, \theta_0, \Theta, \sigma_a^2\}$. Let $Z_T = \{z_1, z_2, \cdots, z_T\}$ an observed time series and assume the initial conditions $Z_* = \{z_{1-p}, \cdots, z_{-1}, z_0\}$ and $a_* = \{a_{1-q}, \cdots, a_{-1}, a_0\}$. The conditional log-likelihood function is

$$\log L_* \left(a_t(\Phi, \theta_0, \Theta); \sigma_a^2\right) = -\frac{T}{2} \log 2\pi\sigma_a^2 - \frac{S_*(\Phi, \theta_0, \Theta)}{2\sigma_a^2} \tag{4}$$

where

$$S_*(\Phi, \theta_0, \Theta) = \sum_{t=1}^{T} a_t^2(\Phi, \theta_0, \Theta | Z_*, a_*, Z_T)$$

is the conditional sum of squares function.

A score function can be obtain as

$$u_* \left(a_t(\Phi, \theta_0, \Theta); \sigma_a^2\right) = \frac{\partial}{\partial \beta} \log L_* \left(a_t(\Phi, \theta_0, \Theta); \sigma_a^2\right) \tag{5}$$

Since the a_t are from a white noise process it is easy to define the Pearson residual from the smooth density

$$f^* \left(a_t(\Phi, \theta_0, \Theta), \hat{F}_T(\Phi, \theta_0, \Theta)\right) = \int k(a_t(\Phi, \theta_0, \Theta); r, g)\, d\hat{F}_T(r; \Phi, \theta_0, \Theta)$$

and the smoothed model

$$m^* \left(a_t(\Phi, \theta_0, \Theta), \sigma_a^2\right) = \int k(a_t(\Phi, \theta_0, \Theta); r, g)\, dM(r; \sigma_a^2)$$

where $\hat{F}_T(r; \Phi, \theta_0, \Theta)$ is the empirical cumulative distribution function based on $a_t(\Phi, \theta_0, \Theta)$, while $M(r; \sigma_a^2)$ is the cumulative distribution function and $k(a_t(\Phi, \theta_0, \Theta); r, g)$ is a kernel density with bandwidth g.

By definition of the Pearson residuals and weight function (Markatou, *et al.* (1998)) we have

$$\delta \left(a_t(\Phi, \theta_0, \Theta); M\left(\sigma_a^2\right), \hat{F}_T(\Phi, \theta_0, \Theta)\right) = \frac{f^* \left(a_t(\Phi, \theta_0, \Theta), \hat{F}_T(\Phi, \theta_0, \Theta)\right)}{m^* \left(a_t(\Phi, \theta_0, \Theta), \sigma_a^2\right)} - 1$$

and

$$w \left(a_t(\Phi, \theta_0, \Theta); M\left(\sigma_a^2\right), \hat{F}_T(\Phi, \theta_0, \Theta)\right) = w(\delta) = \min\left\{1, \frac{[A(\delta)+1]^+}{\delta+1}\right\}$$

where $[\cdot]^+$ indicates the positive part and $A(\cdot)$ is the Residual Adjustment Function, RAF (Lindsay, 1994) that operates on Pearson residuals as the Huber ψ-function operates on the structural residuals. When $A(\delta) = \delta$ the weights are equal to one and this corresponds to maximum likelihood. An example of RAF is $A(\delta) = 2\{(\delta+1)^{1/2} - 1\}$ namely Hellinger RAF that we will use in our examples and simulations. For an extensive discussion of the concept of the RAF see Lindsay (1994).

Hence the Conditional Weighted Likelihood Estimating Equation (cwlee) is

$$\frac{1}{T}\sum_{t=1}^{T} w\left(a_t(\Phi,\theta_0,\Theta); M\left(\sigma_a^2\right), \hat{F}_T\left(r;\Phi,\theta_0,\Theta\right)\right)\ u_*\left(a_t(\Phi,\theta_0,\Theta);\sigma_a^2\right) = 0$$

The likelihood of order m (Azzalini, 1983) approximate the log-likelihood functions by a sum whose generic term is the conditional density function of the corresponding sample element conditional on the m most recent observations, for some $m \geq 0$, i.e.,

$$\log L(\theta) = \sum_{t=1}^{T} \log m(z_t | z_m^{(t-1)};\theta)$$

where $z_m^{(t-1)} = \{z_{t-1},\cdots,z_{t-m}\}$ and $m(z_t|z_m^{(t-1)};\theta)$ is the density function of Z_t given $Z_m^{(t-1)} = z_m^{(t-1)}$.

For an arma model this means to consider the coefficients of the one-step ahead best linear predictor $\Pi(\beta)_m = \{1 - \pi(\beta)_1 B - \cdots - \pi(\beta)_m B^m\}$, $\pi(\beta)_0$ and estimate β from the model:

$$\Pi(\beta)_m Z_t = \pi(\beta)_0 + a_t$$

this leads to the conditional log-likelihood of order m

$$\log L_{*m}\left(a_t(\Pi(\beta)_m);\sigma_a^2\right) = -\frac{T}{2}\log 2\pi\sigma_a^2 - \frac{S_*(\Pi(\beta)_m)}{2\sigma_a^2}$$

that is used in 5 instead of 4 in order to obtain the conditional weighted likelihood estimating equations of order m.

The set of weighted likelihood estimating equations could be solved using the reweighting procedure, i.e., starting with an initial values $\left\{\hat{\beta}_0, \hat{\sigma}_{a0}^2\right\}$ and using the initial conditions Z_* and a_* we can calculate the vector of residuals $\hat{r}_{t0}$. From these residuals the evaluation of the weights are straightforward and leads to solve an estimating equations with fixed weights. With the new values $\left\{\hat{\beta}_1, \hat{\sigma}_{a1}^2\right\}$ the procedure can be iterated until the convergence is achieved. Starting values could be calculated using the bootstrap approach as in Markatou *et al.* (1998). Suppose we have a time series $Y_T = \{y_1, \cdots, y_T\}$ of length T. Fixed the length of subsample to $T_s \leq T$. Then sampling from a uniform discrete distribution in the interval $[T_s, T]$ a value, say t_s and extract the subsample $Y_s = \{y_{t_s}, y_{t_s-1}, \cdots, y_{t_s-T_s+1}\}$ from Y_T. Using Y_s and maximum likelihood estimator we will have one starting value for the reweighting algorithm. Repeating several times the above steps leads to several starting values that gives a way to look for possible multiple roots of the conditional weighted likelihood estimating equations.

Finally, when more than one root arise we can choose the estimator by the parallel disparity measure, see Markatou *et al.* (1998) for details.

4 Robust estimation

In this section we present a procedure based on weighted likelihood and weighted likelihood estimation of order m for robust estimation of the unknown parameter vector β based in three steps: *preliminary estimation*, *outliers classification procedure*, *refined estimation*.

Preliminary estimation: we first estimate β using weighted likelihood when a pure autoregressive model is used and a weighted likelihood estimation of order $m \geq p+q$ in the presence of a moving-average part obtaining the estimates $\hat{\beta}_m$ and the weights $\hat{w}_m(t) = w\left(r_t(\hat{\beta}_m); \hat{\sigma}_a^2, \hat{F}_m(r_t)\right)$ associated to each residual r_t. Using a weighted likelihood estimation of order m we bound the influence of any AO present in the series to at most m observations. Hence with a series of T observations we can expect to have a robust estimation when $[T/2] > n_{AO}\ (m+1) + n_{IO}$, where n_{AO} is the number of additive outliers, n_{IO} is the number of innovations outliers and $[\cdot]$ indicate the integer part.

Outliers classification procedure:

We define a threshold level $0 \leq w_l \leq 1$ in order to decide which observation should be considered as a possible additive outlier. Let $\mathcal{O}_m$ the set contain the observations such that $w_l \geq \hat{w}_m(t)$ and $d_{\mathcal{O}_m}$ its dimension. If $\mathcal{O}_m$ is empty we go to the *refined estimation* step otherwise we try to find the most suitable outliers pattern to describe the present contamination. Each observation in $\mathcal{O}_m$ could be an AO, or not (i.e., an IO or a good observation). All the possible combinations are $f(d_{\mathcal{O}_m}) = 2^{d_{\mathcal{O}_m}}$.

For each combination, namely $\mathcal{O}_m(i)$, $i = 0, \cdots, f(d_{\mathcal{O}_m})$, we consider a modified dataset where we replace every $y_t \in \mathcal{O}_m(i)$ considered as additive outlier with its best linear prediction based on the preliminary estimation β_m, i.e.,

$$y_t = \hat{\pi}(\hat{\beta}_m)_0 + \hat{\pi}(\hat{\beta}_m)_1 y_{t-1} + \cdots + \hat{\pi}(\hat{\beta}_m)_m y_{t-m-1} \tag{6}$$

and the new residual based on the modified sequence

$$\tilde{r}_t = \hat{\Pi}(\hat{\beta}_m)_m y_t - \hat{\pi}(\hat{\beta}_m)_0$$

with $\hat{\Pi}(\hat{\beta}_m)_m = \left(1 - \hat{\pi}(\hat{\beta}_m)_1 B - \cdots - \hat{\pi}(\hat{\beta}_m)_m B^m\right)$. From these $\tilde{r}_t$ we estimate the weights $\hat{w}_{\mathcal{O}_m(i)}(t) = w\left(\tilde{r}_t(\hat{\beta}_m); \hat{\sigma}_a^2, \hat{F}_m(r_t)\right)$. Then we calculate the average weights function

$$\tilde{w}(i) = \frac{\sum_{t \notin \mathcal{O}_m(i)} \hat{w}_{\mathcal{O}_m(i)}(t)}{T} \qquad i = 0, \cdots, f(d_{\mathcal{O}_m}) \tag{7}$$

where we assign weigth zero to each observation in $\mathcal{O}_m(i)$. We choose the outliers pattern $\mathcal{O}_m(i^*)$ such that $\tilde{w}(i^*) = \max_i \tilde{w}(i)$.

We use the average of the weights as a selection criteria since as stated in Markatou *et al.* (1998): "We have found that the sum of the final weights is a useful diagnostic statistic ..., as it tell us roughly how many observations were deleted from the sample".

In this way we hope to bound the influence of any additive outliers on the future residuals and have a robust and more efficient estimation of the unknown parameters.

Refined estimation: After we have classify the outliers we use the corresponding modified dataset defined in 6 based on $\mathcal{O}_m(i^*)$ in order to have a series free of additive outliers so that we can now use a weighted likelihood estimation to obtain a final estimation regardless of the presence of a moving-average part in the model.

Same remark to the introduced outlier classification procedure should be done.

Remark 1: Throughout the steps we will only use $\left\{\hat{\Pi}(\hat{\beta}_m)_m, \hat{\pi}(\hat{\beta}_m)_0\right\}$ and not $\hat{\beta}_m$ so that, for our aims, the weighted likelihood estimation of order m can be replaced by a classical weighted likelihood estimation procedure for an AR(m) model.

Remark 2: A close look should be done to $\tilde{w}(i)$ defined in 7. The kernel density estimation is based on the original time series, i.e., we use the empirical cumulative distribution function $\hat{F}_m(r_t)$ evaluated through the original r_t. Hence all residuals $\tilde{r}_t$ computed without using any modified value will have the same weight. The average of the weights $\tilde{w}(0)$ ($\mathcal{O}_m(0) = \{\emptyset\}$ that is "no one is an AO" is the pattern identified by the preliminary estimation step) corresponds to the usual average of the weights. The co-domain of $\tilde{w}(i)$ is between 0 and 1, on the other hand, $\tilde{w}(i^*) \geq 0.5$ asymptotically, in fact less value will indicate that the major part of the data does not follows the estimated model and are considered as outliers. In $\tilde{w}(i)$, we assign weight zero to every residual associated to a modify y_t, this means that with respect to $\mathcal{O}_m(0)$ we lost its contribution in term of weight. We will choose a different pattern from $\mathcal{O}_m(0)$ only if this lost will be balance by the contribution of the modify y_t will have to the weights of the successive residuals in which y_t appear as an independent variable.

Remark 3: The algorithm is quite fast, since, at most two weighted likelihood estimation equation should be solved and the evaluation of the $\tilde{w}(i)$ is carry out recomputing only few weights each time.

Remark 4: Setting $w_l = 1$ means to look for a possible additive outliers in the whole time series, and this gives the best results. On the other hand, this could make the estimation process unfeasible since the number of all possible combination, hence in practice we choose $w_l < 1$; suitable values should be in the range 0.4-0.6. Further, for $w_l < 1$ an approximated fastest algorithm can be employ, since often it is possible to split the set $\mathcal{O}_m$ in a partition $\{\mathcal{O}_m^{(s)}$ with $s = 1, \cdots, S\}$ where each couple of observations $\{Y_{t_a}^{(s_1)}, Y_{t_b}^{(s_2)}\}$ with $Y_{t_a}^{(s_1)} \in \mathcal{O}_m^{(s_1)}$ and $Y_{t_b}^{(s_2)} \in \mathcal{O}_m^{(s_2)}$ are such that $|t_a - t_b| > m$. In this last case we can run the outliers classification step independently for each $\mathcal{O}_m^{(s)}$ with $s = 1, \cdots, S$ saving in computation time and achieving approximatly the same pattern as we search in $\mathcal{O}_m$.

Remark 5: Note that $\tilde{w}(i)$ is related to the statistics $d = n(1 - \tilde{w}(0))$ introduced in Markatou *et al.* (1998). This statistics can be useful to fix the threshold level w_l since one can evaluated the asymptotic distribution of d under the null hypothesis of no outliers presence and then fix $w_l \leq 1 - d_{1-\alpha}/n$ where $d_{1-\alpha}$ is the $1 - \alpha$ quantile and α is the type one error.

Remark 6: Without contamination the proposed procedure is fully efficient regardless of the w_l choice since the weights will converge uniformly to 1 and the maximum between $\tilde{w}(i)$'s would be in $\tilde{w}(0)$ and no observations will be classified as additive outliers.

Asymptotics properties of the weighted likelihood estimator in the time series setting under the correctly specified model and regularity conditions are equivalent to those of maximum likelihood. In fact the weighted likelihood estimating equations are Fisher consistent, continuous with respect to the unknown parameters and to the distribution function; further the influence function is equal to that of the maximum likelihood, hence by a von Mises expansion we have the results. The weighted likelihood estimation of order m still have the same properties provided that identifiability conditions are satisfy, where $m \geq p + q$ is a necessary condition.

Markatou *et al.* (1998) has show that the weighted likelihood estimating equations as one root with positive breakdown. This is still true in this context provided that no additive outliers are leave in the modified datasets and the number of observations is large enough.

5 Examples

In this section we present two examples. The first one is based on simulated data to the aim of showing the effectiveness of the outliers classification procedure. The second one is based on a real dataset and we perform with it a sensitivity analysis to show the stability of the introduced estimation algorithm. The functions we use are written in R (Ihaka and Gentleman

(1996)) while the simulated time series are generated using fortran ranlib subroutine (www.netlib.org). We use function arima0 of R to perform exact maximum likelihood estimation.

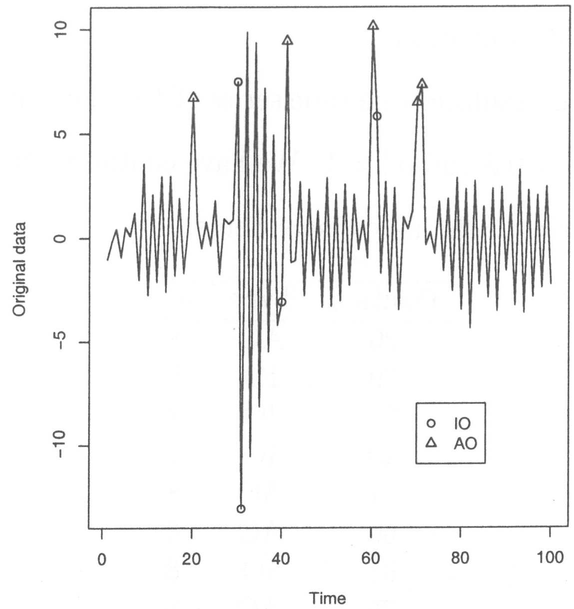
Example: We have simulated one time series of 100 elements from an AR(2) with $\phi_1 = -0.6$, $\phi_2 = 0.3$ and $\sigma_a^2 = 1$. We have contaminated it with several IO and AO outliers with the following scheme:

Position	Type	h_i
20	AO	8
30	IO	8
31	IO	8
40	IO	-8
41	AO	8
60	AO	8
61	IO	-8
70	AO	8
71	AO	8

The contaminated time series is reported in the top part of the figure 1. The contamination level ranging about from $(n_{AO} + n_{IO})/T = 0.09$ to $(3n_{AO} + n_{IO})/T = 0.19$.

Using a conditional weighted likelihood estimating equation we have the refined estimator $\hat{\phi}_1 = -0.683$, $\hat{\phi}_2 = 0.203$, $\hat{\sigma}_a^2 = 0.683$. Preliminary estimation $\hat{\phi}_{1p} = -0.718$, $\hat{\phi}_{2p} = 0.167$, $\hat{\sigma}_{ap}^2 = 0.785$ with threshold $w_l = 0.4$ identify as outliers the observations $20, 21, 30, 31, 40, 41, 42, 43, 60, 61, 70, 71, 72$ while the first ten best pattern are reported in table 1. The $\tilde{w}(0)$ of the preliminary estimation is 0.8257.

Example: In a place call "Col de la Roa" in the Italian Alps there is a meteorological station that record via datalogger several parameters. More than one measure is made every day and, in this example we use the series of mean



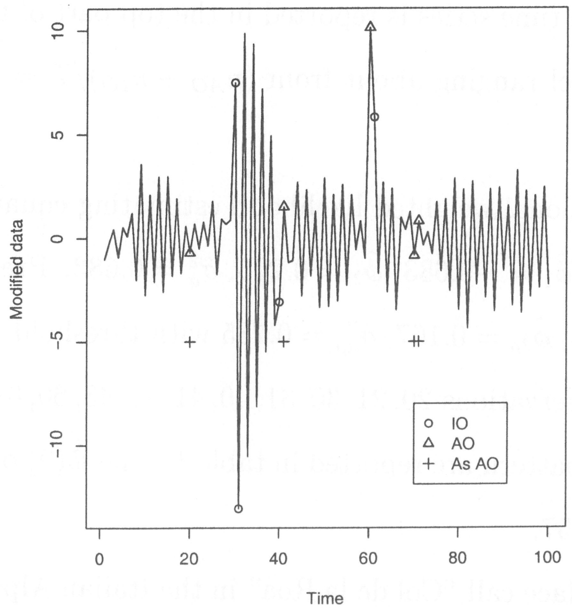


Figure 1: *Up*: The original time series, *Down*: The modified time series.

Obs.	Patterns Is classifies as AO?									
20	x	x	x	x	x	x	x	x		x
21										
30		x								
31										
40			x	x						
41	x	x	x	x	x	x	x	x	x	x
42			x							
43										
60										x
61										
70	x	x	x	x		x		x	x	x
71	x	x	x	x			x	x	x	x
72								x		
$\tilde{w}(i)$	0.8829	0.8802	0.8754	0.8737	0.8736	0.8732	0.8732	0.8731	0.8730	0.8729

Table 1: First ten pattern.

daily temperature (in celsius degrees) for the year 1999 (see figure 2). The original dataset could be download from http://www.tesaf.unipd.it/SanVito/dati.htm.

Since we use only one year, and the data have seasonal behavior we first estimate the seasonal and the trend component using non parametric method as show by the dotted line in figure 2. The resulting residuals are presented in figure 3. This is the series we would like to fit with an arma model.

After we have look to the autocorrelation function and to the partial autocorrelation function (figure 4) we choose an ARMA$(2, 0)$ model. The estimated values with their approximate standard error are reported in table 2 for the exact maximum likelihood estimator and for the conditional weighted likelihood estimator. The results are very similar and the data are well fitted by the model. The new methods highlight two mild outliers, namely observations 98 and 158, that correspond to April 8, and to June 7.

In order to study the stability of the two methods we decide to perform

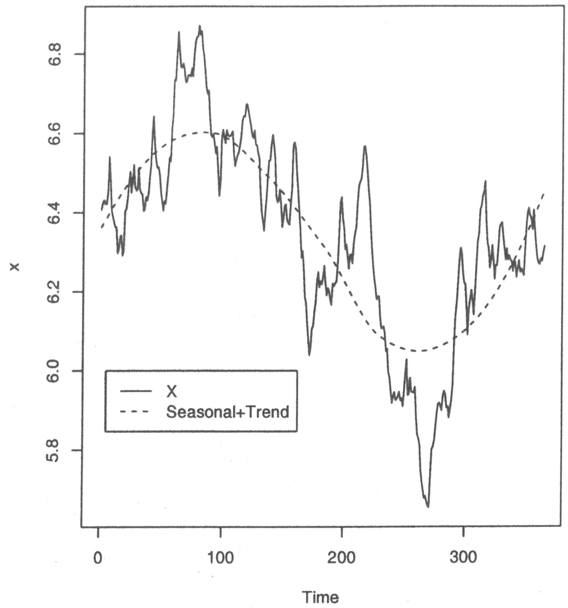


Figure 2: Mean daily temperature for the 1999 from the “Col de la Roa” station (celsius degrees).

	$\hat{\phi}_1$		$\hat{\phi}_2$		$\hat{\sigma}_a^2$
mle	1.314	(0.049)	-0.360	(0.049)	0.00120
cwle	1.314	(0.048)	-0.356	(0.048)	0.00116

Table 2: Estimated values for the “Col del la Roa” dataset (mle = maximum likelihood estimator, cwle = conditional weighted likelihood estimator).

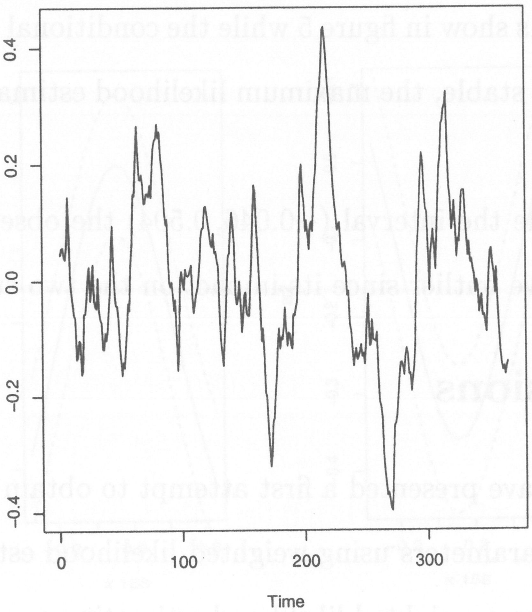


Figure 3: The time series after we remove the seasonal and trend component.

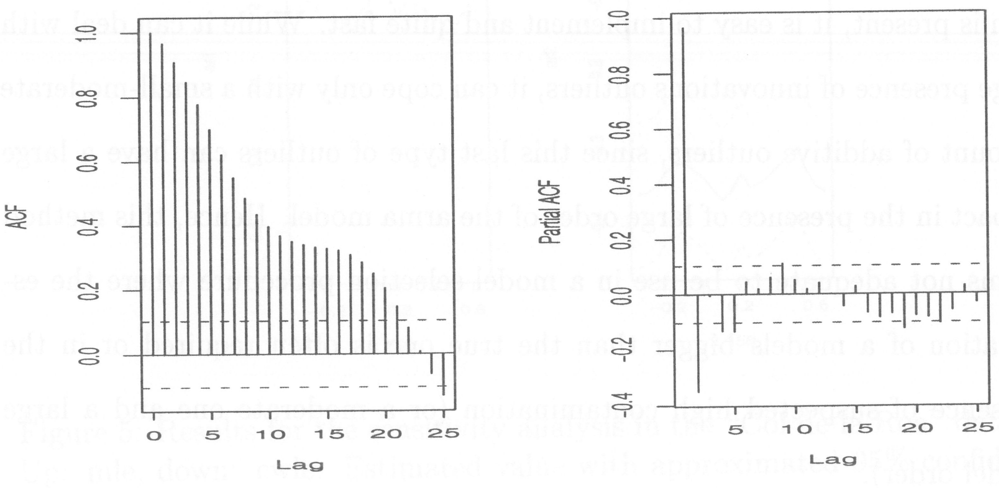


Figure 4: Left: autocorrelation function, right: partial autocorrelation funciton.