Robust Estimation for ARMA models

Nora Muler*, Daniel Peña†

and Víctor J. Yohai‡

Universidad Torcuato di Tella, Universidad Carlos III de Madrid and Universidad de Buenos Aires and CONICET.

November 8, 2007

Abstract

This paper introduces a new class of robust estimates for ARMA models. They are M-estimates, but the residuals are computed so that the effect of one outlier is is limited to the period where it occurs. These estimates are closely related to those based on a robust filter, but they have two important advantages: they are consistent and the asymptotic theory is tractable. We perform a Monte Carlo where we show that these estimates compare favorable with respect to standard M-estimates and to estimates based on a diagnostic procedure.

1 Introduction

There are two main approaches to deal with outliers when estimating ARMA models. The first approach is to start estimating the model parameters using maximum likelihood and then analyzing the residuals with a diagnostic procedure to detect outliers. Among others, diagnostic procedures for detecting outliers were proposed by Fox [10], Chang, Tiao and Chen [5], Tsay [24], Peña [23], and Chen and Liu [6]. However diagnostic procedures suffer from the masking problem: when there are several outliers which have similar effects, the outliers may not be detected.

A second approach is to use robust estimates, that is, estimates which are not much influenced by outlying observations. A detailed review of robust procedures for ARMA models can be found in Chapter 8 of Maronna, Martin and Yohai [17]. In this Chapter it is shown that in the case of an AR(p) model, one outlier at observation t can affect the residuals corresponding to periods t',

*Supported in part by a Grant PAV from ANPCYT.

†Supported in part by Grant SEJ2004-03303 by MEC, Spain.

‡Supported in part by Grant X094 from the University of Buenos Aires, Grant PICT 21407 from ANPCYT Argentina and Grant PIP 5505 from CONICET.

$t \leq t' \leq t+p$; and in the case of an ARMA(p, q) model with $q > 0$ it can affect all residuals corresponding to periods $t' \geq t$. For this reason estimates based on regular residuals (for example M- or S-estimates) are not very robust. One way to improve the robustness of the estimates is to compute the residuals using the robust filter introduced by Masreliez [21]. This robust filter approximates the one step predictor in ARMA models with additive outliers. Several authors have proposed estimates that use residuals computed with the Masreliez filter. For instance, Martin, Samarov and Vandaele [20] proposed filtered M-estimates, Martin and Yohai [19] filtered S-estimates and Bianco, Garcia Ben, Martinez, and Yohai [1] filtered $\tau-$ estimates. However, we can mention two shortcomings of the estimates based on filtered residuals. The first one is that these estimates are asymptotically biased. The second one is that there is not an asymptotic theory for these estimators, and therefore inference procedures like tests or confidence regions are not available.

In this paper we propose a new approach to avoid the propagation of the effect of one outlier when computing the innovation residuals of the ARMA model: we define these residuals using an auxiliary model. For this purpose we introduce the *bounded innovation propagation ARMA* (BIP-ARMA) models. With the help of these models, we are able to define estimates for the ARMA model which are highly robust when the series contains outliers.

We show that the mechanisms of the proposed estimates to avoid the propagation of the outliers are similar to those based on robust filters. However, the advantage of these estimates over those based on the robust filters is that they are consistent and asymptotically normal under a perfectly observed ARMA model.

The proposed estimates can be considered as a generalization of the MM-estimates introduced by Yohai [26] for regression. In the first step we define a highly robust residuals scale and in the second step we use a redescending M-estimate which uses this scale.

For brevity sake we have omitted in this paper some of the proofs. All the proofs can be found in Muler, Peña and Yohai [22].

The paper is organized as follows. In Section 2 we introduce the new family of models and show that the corresponding residuals are similar to those obtained with a robust filter. In Section 3 we introduce the proposed estimates. In Section 4 we establish the main asymptotic results: consistency and asymptotic normality. In Section 5 we discuss the computation of the proposed estimates. In Section 6 we discuss robustness properties of the proposed estimates. In Section 7 we present the results of a Monte Carlo study. In Section 8 we show the performance of the different estimates for fitting a monthly real series. In Section 9 we make some concluding remarks. Section 10 is an Appendix with the main proofs of the asymptotic results.

2 A New Class of Bounded Nonlinear ARMA models

2.1 BIP-ARMA models

We are going to consider a stationary and invertible ARMA model that can be represented by

$$\phi(B)(x_t - \mu) = \theta(B)a_t \tag{1}$$

where a_t are i.i.d. random variables with symmetric distribution and where $\phi(B)$ and $\theta(B)$ are polynomial operators given by $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$ with roots outside the unit circle.

If a_t has first moment we have that $E(x_t) = \mu$. Let $\lambda(B) = \phi^{-1}(B)\theta(B) = 1 + \sum_{i=1}^{\infty} \lambda_i B^i$ and consider the $MA(\infty)$ representation of the ARMA process

$$x_t = \mu + a_t + \sum_{i=1}^{\infty} \lambda_i a_{t-i}. \tag{2}$$

We can model contaminated ARMA processes with a fraction ε of outliers by

$$z_t^{\varepsilon} = (1 - \zeta_t^{\varepsilon})x_t + \zeta_t^{\varepsilon} w_t \tag{3}$$

where y_t is the ARMA model, w_t is an arbitrary process and ζ_t^{ε} is a process taking values 0 and 1 such that $\lim_{n\to\infty} 1/n(\sum_{i=1}^{n} \zeta_t^{\varepsilon}\ ) = \varepsilon$. For example ζ_t^{ε} may be a stationary process such that $E(\zeta_t^{\varepsilon}) = \varepsilon$. The case of additive outliers corresponds to $w_t = x_t + \nu_t$, where x_t and v_t are independent processes. Replacement outliers correspond to the case that the processes x_t and w_t are independent. According to the dependence structure of the process ζ_t^{ε} we can have additive outliers or patchy outliers. For detail, see Martin and Yohai [18]. Robustness is related with the possibility of accurately estimating the parameter of the central model x_t when we observe the contaminated process z_t^{ε}.

Another type of outliers are innovation outliers. An ARMA process with innovation outliers occurs when we observe an ARMA process satisfying (1) but the innovations a_t have a heavy-tailed distribution. Regular M-estimates can cope with this type of outliers. See for example Maronna et al. [17].

We will use the following family of auxiliary models

$$y_t = \mu + a_t + \sum_{i=1}^{\infty} \lambda_i \sigma \eta \left(\frac{a_{t-i}}{\sigma} \right), \tag{4}$$

where the a_t's are i.i.d. random variables with symmetric distribution and σ is a robust M-scale of a_t which coincides with the standard deviation in the case that the a_t 's are normal, the λ_i 's are the coefficients of $\phi^{-1}(B)\theta(B)$ and $\eta(x)$ is an odd and bounded function. An M-scale of a_t is defined as the solution of the equation $E\left(\rho\left(a_t/\sigma\right)\right) = b.$. We call this model, the bounded innovation propagation autoregressive moving average model (BIP-ARMA).

To obtain robust and efficient estimates we will choose η bounded and such that there exists k with $\eta(x) = x$ for $|x| \leq k$. More details on how to choose ρ and b and η are given in Sections 3.1 and 6. Note that in this model the lag effect of a large innovation in period t has a bounded effect on y_{t+j} for any $j \geq 0$ and since $\lambda_j \to 0$ exponentially when $j \to \infty$, this effect will almost disappear in a few periods.

Note that (4) can also be written as

$$y_t = \mu + a_t - \sigma\eta\left(\frac{a_t}{\sigma}\right) + \sigma\phi^{-1}(B)\theta(B)\eta\left(\frac{a_t}{\sigma}\right)$$

and multiplying both sides by $\phi(B)$ we get

$$\phi(B)y_t = \mu(1 - \sum_{i=1}^{p}\phi_i) + \phi(B)a_t - \sigma\phi(B)\eta\left(\frac{a_t}{\sigma}\right) + \sigma\ \theta(B)\eta\left(\frac{a_t}{\sigma}\right)$$

which is equivalent to

$$y_t = a_t + \mu + \sum_{i=1}^{p}\phi_i(y_{t-i} - \mu)\ - \sum_{i=1}^{r}\left(\phi_i a_{t-i} + (\theta_i - \phi_i)\sigma\eta\left(\frac{a_{t-i}}{\sigma}\right)\right), \qquad (5)$$

where $r = \max(p, q)$. If $r > p$, $\phi_{p+1} = \cdots = \phi_r = 0$ and if $r > q$, $\theta_{q+1} = \cdots = \theta_r = 0$.

2.2 Robust filters and BIP-ARMA models

Let us analyze the relationship of the BIP-ARMA model and an ARMA model with additive outliers. The BIP-ARMA model can be also be written as $y_t = (1-\zeta_t^{\varepsilon})x_t + \zeta_t^{\varepsilon}(x_t + \nu_t)$, where $x_t = y_t - a_t + a_t^*$ is an ARMA model satisfying $\phi(B)(x_t - \mu) = \theta(B)a_t^*$, $a_t^* = \sigma\eta(a_t/\sigma)$, $\nu_t = a_t - a_t^*$, $\zeta_t^{\varepsilon} = I(|a_t| \geq k)$ and $\varepsilon = P((|a_t| \geq k)$. However, in the BIP-ARMA model ζ_t^{ε} and ν_t are not independent and they are also not independent of x_t.

We will show that the one-step forecast in the BIP-ARMA model is similar to the forecast obtained by using the robust filter for ARMA models introduced by Masreliez [21]. The Masreliez filter was proposed as an approximation to one-step predictor for additive models of the form (3), where x_t is a Gaussian ARMA model, ζ_t^{ε} are i.i.d. Bernoulli variables with $P(\zeta_t^{\varepsilon} = 1) = \varepsilon$ and ν_t are i.i.d. normal random variables.

Suppose that we have an ARMA series $y_1, ..., y_n$ and that we suspect that it is contaminated with additive outliers. Assume first that we know the parameters $\boldsymbol{\phi}, \boldsymbol{\theta}, \mu$ and σ of the ARMA model. The robust filter computes a "clean"series y_t^*, and filtered innovations residuals $\widehat{a}_t$ that are obtained by the following recursive procedure. Suppose that the cleaned series $y_1^*, ..., y_{t-1}^*$, and the filtered innovation residuals $\widehat{a}_1, ..., \widehat{a}_{t-1}$ previous to time t are computed. Since $y_t = \mu - \sum_{i=1}^{\infty}\pi_i(y_{t-i} - \mu)\ + a_t$, where $\pi(B) = \theta(B)^{-1}\phi(B) = 1 + \sum_{i=1}^{\infty}\pi_i B^i$, the one-step ahead robust forecast of y_t, is obtained by replacing the y_{t-i}'s by the cleaned values y^*_{t-i} 's, i.e., the one-step robust forecast of y_t is obtained by

$$\widehat{y}^*_t = \mu - \sum_{i=1}^{\infty} \pi_i (y^*_{t-i} - \mu) = \mu - (\theta(B)^{-1}\phi(B) - 1)(y^*_t - \mu), \tag{6}$$

where $y^*_t = \mu$ for $t \leq 0$. The filtered innovation residual for period t is computed by $\widehat{a}^*_t = y_t - \widehat{y}^*_t$ and the cleaned value y^*_t by

$$y^*_t = \widehat{y}^*_t + s_t \eta^* \left(\frac{\widehat{a}^*_t}{s_t} \right) = y_t - \widehat{a}^*_t + s_t \eta^* \left(\frac{\widehat{a}^*_t}{s_t} \right), \tag{7}$$

where s_t is an estimate of the one-step prediction error scale and where η^* has the same properties as those stated for η, in particular for some $k > 0$ it holds that $\eta^*(u) = u$ for $|u| \leq k$. Observe that if $|\widehat{a}^*_t| \leq k$, then $s_t \eta^* (\widehat{a}^*_t / s_t) = \widehat{a}^*_t$ and $y^*_t = y_t$. Recursive formulae for s_t can be found in Martin et al. [20].

We can easily derive from (6) and (7) that

$$\widehat{y}^*_t = \mu + \sum_{i=1}^{p} \phi_i (y^*_{t-i} - \mu) - \sum_{i=1}^{q} \theta_i s_t \eta^* \left(\frac{\widehat{a}^*_{t-i}}{s_t} \right). \tag{8}$$

Now, from (5), the one step forecast for y_t in the BIP-ARMA model is given by

$$\widehat{y}_t = \mu + \sum_{i=1}^{p} \phi_i \left(y_{t-i} - \mu - a_{t-i} + \sigma \eta \left(\frac{a_{t-i}}{\sigma} \right) \right) - \sum_{i=1}^{q} \theta_i \sigma \eta \left(\frac{a_{t-i}}{\sigma} \right), \tag{9}$$

which is similar to (8) taking as the cleaned series

$$y^*_t = y_t - a_t + \sigma \eta \left(a_t / \sigma \right). \tag{10}$$

The main difference is that here s_t is taken is taken constant and equal to σ. Thus, the filtered residuals used by Martin et al. [20] and Bianco et al. [1] are very similar to those of a BIP-ARMA model. In the next Section we will use the model (4) to define robust estimates of the parameters of an ARMA model that may contain additive outliers.

3 Bounded MM-estimates for ARMA models

Assume that $y_1, ..., y_n$ are observations corresponding to a BIP-ARMA model and that the density of a_t is $f(u)$. The conditional log likelihood function of $y_{p+1}, ..., y_n$ given $y_1, ..., y_p$ and the values $a^b_{p-r+1}(\boldsymbol{\beta}, \sigma) = 0, ..., a^b_p(\boldsymbol{\beta}, \sigma) = 0$, where $r = \max(p, q)$, can be written as

$$L_c(\boldsymbol{\beta}, \sigma) = \sum_{t=p+1}^{n} \log f \left(a^b_t(\boldsymbol{\beta}, \sigma) \right), \tag{11}$$

where from (5), the functions $a_t^b(\boldsymbol{\beta},\sigma)$ are defined recursively for $t \geq p+1$ by

$$a_t^b(\boldsymbol{\beta},\sigma) = y_t - \mu - \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{i=1}^{r} \left(\phi_i a_{t-i}^b(\boldsymbol{\beta},\sigma) + (\theta_i - \phi_i)\, \sigma \eta \left(\frac{a_{t-i}^b(\boldsymbol{\beta},\sigma)}{\sigma} \right) \right). \tag{12}$$

In the case of a pure ARMA model, i.e., where $\eta(u) = u$, (12) reduces to

$$a_t(\boldsymbol{\beta}) = y_t - \mu - \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}(\boldsymbol{\beta}). \tag{13}$$

Since ML-estimates are not robust , we will consider M-estimates which minimizes

$$M_n^b(\boldsymbol{\beta}) = \frac{1}{n-p} \sum_{t=p+1}^{n} \rho \left(\frac{a_t^b(\boldsymbol{\beta},\widehat{\sigma})}{\widehat{\sigma}} \right), \tag{14}$$

where ρ is a bounded function, and $\widehat{\sigma}$ is an estimate of σ.

We observe that the M-estimates defined in (14), require to have an estimate $\widehat{\sigma}$ of σ. This lead us to define in Section 3.2 a two step procedure for estimating $\boldsymbol{\beta}$ that we call MM-estimates.

3.1 M-estimates of scale

Huber [13] introduced the M-estimates of scale. Given a sample $\mathbf{u} = (u_1, ..., u_n)$, $u_i \in R$, an M-estimate of scale $S_n(\mathbf{u})$ is defined by any value $s \in (0, \infty)$ satisfying

$$\frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{u_i}{s} \right) = b, \tag{15}$$

where ρ is a function satisfying the following property P1 given below.

P1: $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ is continuous, non-constant and non-decreasing in $|x|$.

We can define two asymptotic breakdown points of an M-estimate of scale: the minimum fraction of outliers which are required to make the estimate infinity, ϵ_∞^*, and the minimum fraction of inliers that may take the estimate to zero, ϵ_0^*. Huber [14] shows that $\epsilon_\infty^* = b/a$ and $\epsilon_0^* = 1 - b/a$ where $a = \max \rho$. Then, the global breakdown point of these estimates is $\epsilon^* = \min(\epsilon_\infty^*, 1 - \epsilon_\infty^*)$ and so taking $b = a/2$ we get a maximum breakdown point of 0.5. To make the M-scale estimate consistent for the standard deviation when the data are normal, we require that $E_\Phi(\rho(x)) = b$ where Φ is the standard normal distribution.

3.2 MM-Estimates

The MM-estimates for regression were introduced by Yohai [26] to combine high breakdown point with high efficiency under normal errors. The key idea of the MM-estimates is to compute in the first step a highly robust estimate of the error scale, and in the second step this scale estimate is used to compute an

M-estimate of the regression parameters. For time series models these two steps are not enough to guarantee robustness. This is due to the fact that an a outlier in one period, does not only affect the residual corresponding to this period, but it may also affect all the subsequent residuals. To avoid this propagation we define MM-estimates for the ARMA model, where the residuals are computed as in the BIP-ARMA model instead as in the regular ARMA model. Then, the procedure for computing MM-estimates is as follows.

Step 1. In this step we obtain an estimate of σ. For this purpose we consider two estimates of σ, one using an ARMA model and another using a BIP-ARMA model and choose the smallest one.

Let ρ_1 be a bounded function satisfying P1 and such that if $b = E(\rho_1(u))$, then $b/\max\rho_1(u) = 0.5$. This guarantees that for a normal random sample the M-scale estimator s based on ρ_1 converges to the standard deviation and that the breakdown point of s is 0.5. Put

$$\mathcal{B}_{0,\zeta} = \left\{(\boldsymbol{\phi},\boldsymbol{\theta}) \in R^{p+q}\colon\ |z| \geq 1+\zeta \text{ holds for all the roots } z \text{ of } \phi(B) \text{ and } \theta(B)\right\}. \tag{16}$$

Let us call $\mathcal{B}_0 = \mathcal{B}_{0,\zeta}$ for some small $\zeta > 0$ and $\mathcal{B} = \mathcal{B}_{0,\zeta} \times R$. Then, we define an estimate of $\boldsymbol{\beta}$

$$\widehat{\boldsymbol{\beta}}_S = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} S_n(\mathbf{a}_n(\boldsymbol{\beta})) \tag{17}$$

and the corresponding estimate of σ

$$s_n = S_n(\mathbf{a}_n(\widehat{\boldsymbol{\beta}}_S)), \tag{18}$$

where $\mathbf{a}_n(\boldsymbol{\beta}) = (a_{p+1}(\boldsymbol{\beta}),...,a_n(\boldsymbol{\beta}))$ and S_n is the M-estimate of scale based on ρ_1 and b.

Let us describe now the estimate corresponding to the BIP-ARMA model. Define $\widehat{\boldsymbol{\beta}}_S^b = (\widehat{\boldsymbol{\phi}}_S^b,\widehat{\boldsymbol{\theta}}_S^b,\widehat{\mu}_s^b)$ by the minimization of $S_n(\mathbf{a}_n^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})))$ over $\mathcal{B}$. The value $\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})$ is an estimate of σ computed as if $(\boldsymbol{\phi},\boldsymbol{\theta})$ were the true parameters and the a_t's were normal. Then since in this case the M-scale σ coincides with the standard deviation of a_t, from (4) we have

$$\sigma^2 = \frac{\sigma_y^2}{\left(1+\kappa^2\sum_{i=1}^{\infty}\lambda_i^2(\boldsymbol{\phi},\boldsymbol{\theta})\right)},$$

where κ^2 =var$(\eta(a_t/\sigma))$ and σ_y^2 =var(y_t). Let $\widehat{\sigma}_y^2$ be a robust estimate of σ_y^2 and κ^2 =var$(\eta(z))$ where z has N(0,1) distribution. Then, we define

$$\widehat{\sigma}^2(\boldsymbol{\phi},\boldsymbol{\theta}) = \frac{\widehat{\sigma}_y^2}{\left(1+\kappa^2\sum_{i=1}^{\infty}\lambda_i^2(\boldsymbol{\phi},\boldsymbol{\theta})\right)}. \tag{19}$$

The scale estimate s_n^b corresponding to the BIP-ARMA model is defined by

$$\widehat{\boldsymbol{\beta}}_S^b = (\widehat{\boldsymbol{\phi}}_S^b,\widehat{\boldsymbol{\theta}}_S^b,\widehat{\mu}_s^b) = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} S_n(\mathbf{a}_n^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))) \tag{20}$$

and

$$s_n^b = S_n\left(\mathbf{a}_n^b\left(\widehat{\boldsymbol{\beta}}_S^b, \widehat{\sigma}(\widehat{\boldsymbol{\phi}}_S^b, \widehat{\boldsymbol{\theta}}_S^b)\right)\right), \tag{21}$$

where $\mathbf{a}_n^b(\boldsymbol{\beta}, \sigma) = (a_{p+1}^b(\boldsymbol{\beta}, \sigma), ..., a_n^b(\boldsymbol{\beta}, \sigma))$. Our estimate of σ is

$$s_n^* = \min(s_n, s_n^b). \tag{22}$$

As we will see in the next section, if the sample is taken from an ARMA model without outliers, asymptotically we obtain $s_n < s_n^b$. We should point out that despite the fact that κ was computed as if the a_t's were normal, the asymptotic properties of the estimators are not going to depend on this assumption.

Step 2. Consider a bounded function ρ_2 such that satisfies P1 and $\rho_2 \leq \rho_1$. This function is chosen so that the corresponding M-estimate is highly efficient under normal innovations. Let

$$M_n(\boldsymbol{\beta}) = \frac{1}{n-p} \sum_{t=p+1}^{n} \rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right) \tag{23}$$

and

$$M_n^b(\boldsymbol{\beta}) = \frac{1}{n-p} \sum_{t=p+1}^{n} \rho_2\left(\frac{a_t^b(\boldsymbol{\beta}, s_n^*)}{s_n^*}\right). \tag{24}$$

We define the estimates $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\beta}}_M^b$ by the minimization over $\mathcal{B}$ of $M_n(\boldsymbol{\beta})$ and $M_n^b(\boldsymbol{\beta})$ respectively. Then, the MM-estimate $\widehat{\boldsymbol{\beta}}_M^*$ is equal to $\widehat{\boldsymbol{\beta}}_M$ if $M_n(\widehat{\boldsymbol{\beta}}_M) \leq M_n^b(\widehat{\boldsymbol{\beta}}_M^b)$ and is equal to $\widehat{\boldsymbol{\beta}}_M^b$ if $M_n(\widehat{\boldsymbol{\beta}}_M) > M_n^b(\widehat{\boldsymbol{\beta}}_M^b)$.

For instance we can take $\rho_2(x) = \rho_1(\lambda x)$ with $0 < \lambda < 1$. If $\rho_2''(0) > 0$, ρ_2 will be close to a quadratic function when λ tends to 0.

Remark 1. One important problem that will be only briefly mentioned here is that of the robust model selection. One possibility to explore is to adapt to ARMA models the robust finite prediction error (RFPE) selection criterion given in Section 5.12 of Maronna et al. [17] for regression .

In the next section we will show that when the sample is taken from an ARMA model without outliers, for large n the estimate will choose $\widehat{\boldsymbol{\beta}}_M^* = \widehat{\boldsymbol{\beta}}_M$; and in our Monte Carlo study of Section 7 we observe that if the sample has enough additive outliers we may have $\widehat{\boldsymbol{\beta}}_M^* = \widehat{\boldsymbol{\beta}}_M^b$. This implies that $\widehat{\boldsymbol{\beta}}_M^*$ and $\widehat{\boldsymbol{\beta}}_M^b$ have the same asymptotic distribution for any η. However, the efficiency of $\widehat{\boldsymbol{\beta}}_M^b$ for finite sample size depends of η. If the interval where η coincides with the identity increases, the efficiency for finite sample size of $\widehat{\boldsymbol{\beta}}_M^b$ will increase too, but the propagation of the outliers effect will gain importance and so the estimate will lose robustness.

4 Asymptotic results

The main results of this Section, stated in Theorems 4 and 6, are the consistency and asymptotic normality of the BMM-estimators for ARMA models. These theorems require to prove first the consistency of S- and the consistency and asymptotic normality of MM-estimators. We stated these results in Theorems 1, 3 and 5 respectively. The link that relates the properties of S- and MM- to those of BMM-estimates are Theorems 2 and 4.

Consider the following assumptions:

P2. The process y_t is a stationary and invertible ARMA (p, q) process with parameter $\boldsymbol{\beta}_0 = (\boldsymbol{\phi}_0, \boldsymbol{\theta}_0, \mu_0) \in \mathcal{B}$ and $E(\log^+ |a_t|) < \infty$, where $\log^+ a = \max(\log a, 0)$. The polynomials $\phi_0(B)$ and $\theta_0(B)$ do not have common roots.

P3. The innovation a_t has an absolutely continuous distribution with a symmetric and strictly unimodal density.

P4. $P(a_t \in C) < 1$ for any compact C.

P5. The function η is continuous, even and bounded.

The following Theorem establishes the consistency of the S-estimates based on ARMA models.

Theorem 1. *Assume that y_t satisfies P2 with innovations a_t satisfying P3. Assume also that ρ_1 is bounded and satisfies P1 with $\sup \rho_1 > b$, and that $\psi_1 = \rho_1'$ is bounded and continuous. Then, (i) The estimate $\widehat{\boldsymbol{\beta}}_S$ defined in (17) is strongly consistent for $\boldsymbol{\beta}_0$. (ii)*

Let s_n be the scale estimate defined in (18). Then $s_n \longrightarrow s_0$ a.s. where s_0 is defined by $E\left(\rho_1\left(a_t/s_0\right)\right) = b$.

The next Theorem establishes that under a regular ARMA model $\widehat{\boldsymbol{\beta}}_S$ and $\widehat{\boldsymbol{\beta}}_S^b$ are asymptotically equivalent.

Theorem 2. *Assume that y_t satisfies condition P2, with innovations a_t satisfying P3 and P4. Assume also that ρ_1 is bounded and satisfies P1 with* $\sup \rho_1 > b$, *that $\psi_1 = \rho_1'$ is bounded, continuous and that η satisfies P5. Then if y_t is not white noise, with probability 1 there exists n_0 such that $\widehat{\boldsymbol{\beta}}_S^b = \widehat{\boldsymbol{\beta}}_S$ for all $n \geq n_0$ and then s_n^* defined in (22) verifies $s_n^* \longrightarrow s_0$ a.s..*

The reason why the above theorem requires that y_t is not white noise is that in that case both models: the regular ARMA and the BIP-ARMA with any function η, coincides. Therefore, in this case it does not matter which of the two model is chosen.

The following Theorems shows the consistency of the MM-estimate.

Theorem 3. *Assume that y_t satisfies condition P2, with innovations a_t satisfying P3. Assume also that ρ_i , i=1,2, are bounded and satisfy P1, $\psi_i = \rho_i'$, i=1,2 are bounded and continuous and that* $\sup \rho_1 > b$. *Then $\widehat{\boldsymbol{\beta}}_M \longrightarrow \boldsymbol{\beta}_0$ a.s..*

The next Theorem shows that asymptotically under a regular ARMA model $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\beta}}_M^b$ are equivalent.

Theorem 4. *Suppose that the assumptions of Theorem 3 , P4 and P5 hold. Then if y_t is not white noise, with probability 1 there exists n_0 such that $\widehat{\boldsymbol{\beta}}_M^b = \widehat{\boldsymbol{\beta}}_M$ for all $n \geq n_0$ and then $\widehat{\boldsymbol{\beta}}_M^* \rightarrow \boldsymbol{\beta}_0$ a.s..*

The following Theorem shows the asymptotic normality of the MM-estimates.

Theorem 5. *Suppose that the assumptions of Theorem 3 hold. Moreover assume that ψ_2' and ψ_2'' are continuous and bounded functions and $\sigma_a^2 = E(a_t^2) < \infty$. Then we have*

$$(n-p)^{1/2}(\widehat{\boldsymbol{\beta}}^M - \boldsymbol{\beta}_{\mathbf{0}}) \ \rightarrow_D N(\mathbf{0}, D),$$

where

$$D = \frac{s_0^2 E_{F_0}\left(\psi_2^2\left(a_t/s_0\right)\right)}{E_{F_0}^2(\psi_2'\left(a_t/s_0\right))} \begin{pmatrix} \sigma_a^{-2}\mathbf{C}^{-1} & \mathbf{0} \\ \mathbf{0} & \zeta_0^{-2} \end{pmatrix}, \tag{25}$$

$$\zeta_0 = -\frac{1 - \sum_{i=1}^{p} \phi_{0i}}{1 - \sum_{i=1}^{q} \theta_{0i}}$$

and $C = (c_{ij})$ is the $(p+q+1) \times (p+q+1)$ matrix given by

$$c_{i,j} = \sum_{k=0}^{\infty} \upsilon_k \upsilon_{k+j-i} \quad \text{if} \quad i \leq j \leq p$$

$$c_{p+i,p+j} = \sum_{k=0}^{\infty} \varpi_k \varpi_{k+j-i} \quad \text{if} \quad i \leq j \leq q,$$

$$c_{i,p+j}=-\sum_{k=0}^{\infty}\varpi_k\upsilon_{k+j-i}\quad\text{if}\quad i\leq p,\ j\leq q, i\leq j$$

$$c_{i,p+j}=-\sum_{k=0}^{\infty}\upsilon_k\varpi_{k+i-j}\quad\text{if}\quad i\leq p,\ j\leq q,\ j\leq i\ ,$$

where $\phi_0^{-1}(B)=1+\sum_{i=1}^{\infty}\upsilon_iB^i$ and $\theta_0^{-1}(B)=1+\sum_{i=1}^{\infty}\varpi_iB^i$. Observe that when the a_t's are normal, $\sigma^2=\sigma_a^2$.

Remark 2. When $\rho_2(u)=u^2$, $\widehat{\boldsymbol{\beta}}_M$ is the conditional maximum likelihood estimate corresponding to normal errors. Let F_0 be the distribution of a_t, then, in this case $s_0^2E_{F_0}\left(\psi_2^2\left(a_t/s_0\right)\right)/E_{F_0}^2(\psi_2'\left(a_t/s_0\right))\ =\sigma_a^2$ Therefore the asymptotic efficiency of the MM-estimate with respect to the normal conditional maximum likelihood estimate when the innovations have distribution F_0 is

$$\mathrm{EFF}\left(\psi_2,F_0\right)=\frac{\sigma_a^2E_{F_0}^2(\psi_2'\left(a_t/s_0\right))}{s_0^2E_{F_0}\left(\psi_2^2\left(a_t/s_0\right)\right)}. \tag{26}$$

Choosing ψ_2 conveniently we can make this efficiency as close to one as desired for the case that F_0 is normal.

Remark 3. The relative efficiency of the MM- and BMM- estimates given by (26) is the same than the one of the M-estimates of location with respect to the mean. This implies the well-know fact that M-estimates are robust for innovation outliers, that is when $y_t, 1\leq t\leq n$, correspond to a perfectly observed ARMA model, but the distribution F_0 of a_t is heavy tailed.

Remark 4. When $E(a_t^2)=\infty$, the rate of convergence of M-estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ may be larger than $n^{-1/2}$, and the asymptotic distribution non-normal. This case was studied by several authors. See for example Davis, Knight and Liu [8] and Davis [9].

Remark 5. Theorems 1-5 use P3 only to guarantee that for all σ, the function $g(\mu)$= $E(\rho((a_t-\mu)/\sigma)$ has a unique minimum at $\mu=0$. If $g(\mu)$ has a unique minimum at $\overline{\mu}\neq 0$, then the estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are still consistent, but the estimate of μ will converge to $\mu_0+\overline{\mu}$.

Finally from Theorems 4 and 5 we derive the following Theorem.

Theorem 6. *Suppose that the assumptions of Theorem 5, P4 and P5 hold. Then* $(n-p)^{1/2}(\widehat{\boldsymbol{\beta}}_M^{*}-\beta_0)$ *converges in distribution to a* $N(0,D)$ *distribution, where* D *is defined in (25).*

Note that the assumptions of Theorems 2 and 4 include condition P4. However this condition is not strictly necessary and is included only to simplify the proofs.

All the asymptotic theorems of this Section assume that the process is an ARMA model. We conjecture that similar results, consistency and asymptotic normality hold when the observations follows a BIP-ARMA model. The main difficulty to prove these results is to show that the distribution of $a_t^b(\boldsymbol{\beta},\sigma)$ is asymptotically stationary.

5 Computation

We will discuss here how to compute the MM-estimate. We start computing the estimates of step 1, $\widehat{\boldsymbol{\beta}}_S$ and $\widehat{\boldsymbol{\beta}}_S^b$. According to (15) we can write $S_n^2(\mathbf{a}_n(\boldsymbol{\beta})\ ) = \sum_{t=p+1}^n r_t^2(\boldsymbol{\beta})$, where

$$r_t(\boldsymbol{\beta}) = \frac{S_n(\mathbf{a}_n(\boldsymbol{\beta})\ )}{(n-p)^{1/2}\, b^{1/2}}\ \rho_1^{1/2}\left(\frac{a_t(\boldsymbol{\beta})}{S_n(\mathbf{a}_n(\boldsymbol{\beta}))}\right).$$

Then to compute $\widehat{\boldsymbol{\beta}}_S$ we can use any non-linear least squares algorithm, for example a Marquard algorithm. Similarly we can transform the minimization of $S_n(\mathbf{a}_n^b(\boldsymbol{\beta},\widehat{\sigma}\,(\boldsymbol{\phi},\boldsymbol{\theta})\,))$ in a non-linear least squares problem. Note that non-linear least squares algorithms require a good starting point. Since the functions we are minimizing are non-convex and they may have several local minima, the choice of the starting point is crucial.

If the model has few parameters (e.g., $p+q \leq 3$), one way to obtain the starting point is to generate a grid of values of the parameter and choose as initial estimate the one minimizing the objective function. Note that the case of $p+q \leq 3$ is very frequent in the case of ARMA applications, where the use of parsimonious models is recommended. Bianco et al. [1] gave an algorithm to compute a highly robust starting point when there are more parameters.

In the second step, to compute $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\beta}}_M^b$ we can use Marquard algorithm using a similar idea and taking as initial estimate the best estimate computed in step 1.

In our simulations the estimates were defined taking

$$\rho_2\,(x) = \begin{cases} 0.5x^2 & \text{if} \quad |x| \leq 2 \\ 0.002x^8 - 0.052x^6 + 0.432x^4 - 0.972x^2 + 1.792 & \text{if} \quad 2 < |x| \leq 3 \\ 3.25 & \text{if} \quad |x| > 3, \end{cases}$$

$\rho_1\,(x) = \rho_2\,(x/0.405)$ and $\eta = \rho_2'$.Note that ρ_1 and ρ_2 are smooth functions which are quadratic in the intervals $(-0.81, 0.81)$ and $(-2, 2)$ respectively. The function ρ_1 was chosen so that if we take $b = \max \rho_1/2$ then the scale is consistent to the standard deviation for normal samples Note that η is a redescending function.

For fitting an ARMA(1,1) model to 1000 observations using a MATLAB program, with an initial solution computed with a grid of 20 values in each parameter, the computing time of a BMM-estimate with the choices of ρ_i, $i = 1, 2$ and η given above is approximately 10 seconds in a PC computer with an AMD Athlon 1.8 GHz processor. For fitting an AR(3) model under the same conditions, the computing time is 1 minute 20 seconds.

6 Robustness properties

Several robustness measures can be used for estimates of time series parameters. Hampel [12] introduced the influence curve to measure the robustness of an estimate under an infinitesimal outlier contamination in the framework of i.i.d. observations. Künsch [15], Martin and Yohai [18] and Mancini, Ronchetti and Trojani [16] give generalizations of the influence curve for estimating time series parameters. However, because of its infinitesimal character, the influence curve may not be a good measure of the robustness when there is a positive fraction of outlier contamination. For example, it can be proved that for a very small amount of contamination the MM- and BMM-estimates asymptotically coincide and therefore their influence curves also coincide. However, we will see below in this Section and in Section 7 that the BMM-estimate is more robust than the MM-estimate. Influence functions for the M-estimates of ARMA models can be found in Martin and Yohai [18].

A more reliable measure of the robustness of an estimate to cope with a positive fraction ε of contamination is the asymptotic maximum bias. Consider a family of $\varepsilon-$contaminated process

$$z_t^{\varepsilon k} = (1-\zeta_t^{\varepsilon})x_t + \zeta_t^{\varepsilon} w_t^k \tag{27}$$

as in (3) where $k \in K$ and $(x_t, \zeta_t^{\varepsilon}, w_t^k)$ is stationary. Suppose also that the distribution of the uncontaminated process y_t depends on a parameter $\boldsymbol{\gamma} \in \Gamma \subset R^j$. As example we can consider the family of additive outliers models which is obtained taking $w_t^k = x_t + k$, with $k \in R$

Suppose that for each sample size n we have an estimate $\widehat{\boldsymbol{\gamma}}_n$ of $\boldsymbol{\gamma}$ and let $\widehat{\boldsymbol{\gamma}}_\infty(L) = (\widehat{\gamma}_{\infty 1}(L), ..., \widehat{\gamma}_{\infty j}(L))$ be the almost sure limit of $\widehat{\boldsymbol{\gamma}}_n$ when applied to a process with distribution L. The bias of the i-th component $\widehat{\boldsymbol{\gamma}}_\infty$when applied to $z_t^{\varepsilon k}$ as defined in (27) is

$$B(\widehat{\gamma}_{\infty i}, \gamma, \varepsilon, k) = |\widehat{\gamma}_{\infty i}(\mathcal{L}(z_t^{\varepsilon k})) \ - \gamma_i|,$$

where $\mathcal{L}(z_t^{\varepsilon k})$ denotes de distribution of the process $z_t^{\varepsilon k}$. The maximum asymptotic bias of the i-th component is defined by

$$MB(\widehat{\gamma}_{\infty i}, \gamma, \varepsilon) = \sup_{k \in K} \ B(\widehat{\gamma}_{\infty i}, \gamma, \varepsilon, k).$$

We have approximately computed the maximum bias curves of the MM- and BMM-estimates for Gaussian AR(1) and MA(1) models with additive outliers ($w_t^k = x_t + k$) and where the ζ_t^{ε} are i.i.d. Bernoulli variables. To simplify the computation we eliminate the intercept from these models by assuming it to be known and null. The asymptotic value of the estimate is approximated using samples of size 10000. We found that for samples size larger than 10000 the changes in the estimate are negligible.

In Figure 1 we show the bias curves of the MM- and BMM-estimates for the AR(1) model with $\phi = 0.5$ and $\varepsilon = 0.1$. In Figure 2 we show the maximum biases curves for the MM- and BMM-estimates under the same model In Figure 3 we show the maximum bias curve for the BMM-estimate under a MA(1) model with parameter $\theta = -0.5$.

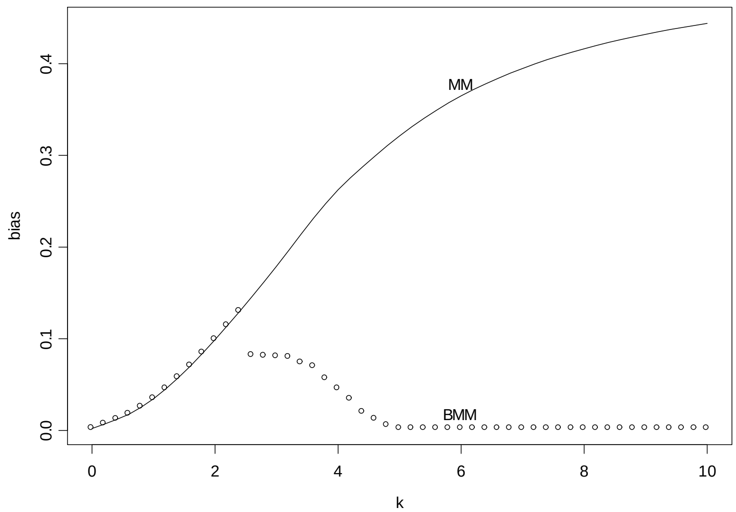

Figure 1: Bias curve for the AR(1) model with $\phi = .5$ and 10% of additive outliers where k is the outlier size

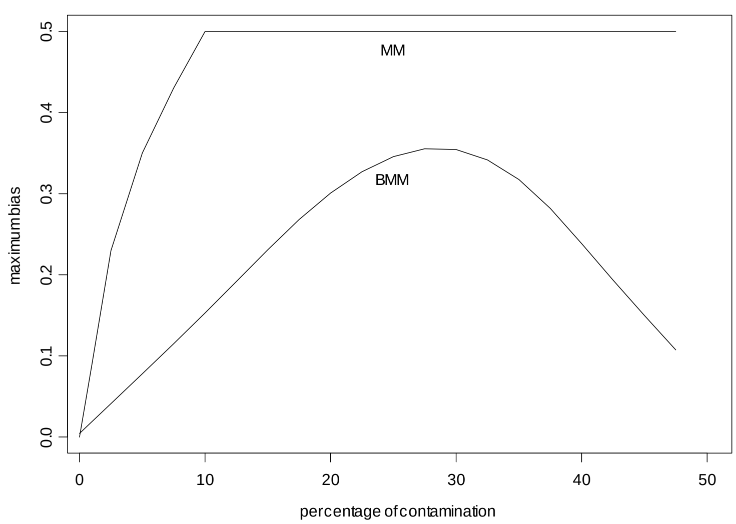

Figure 2: Maximum Bias for the AR(1) model with $\phi = .5$

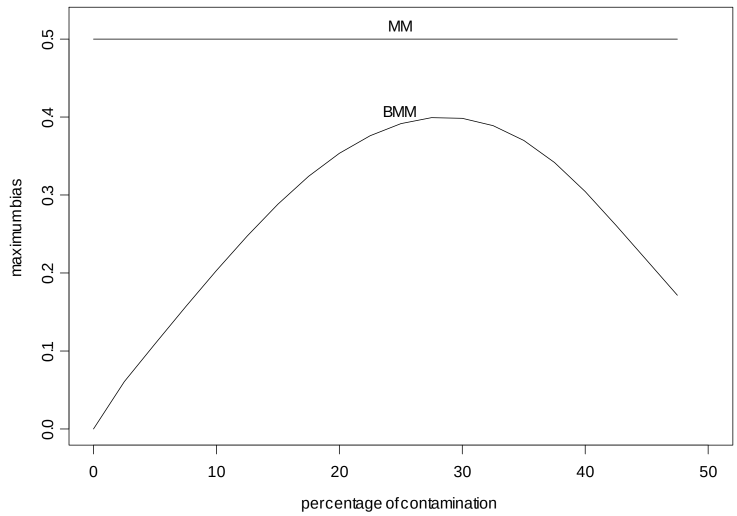

Figure 3: Maximum Bias of the BMM for the MA(1) model with $\theta = -.5$

In both cases we observe that the BMM-estimate has a smaller maximum bias than the MM-estimate. We also observe that the behavior of the MM- is different from the BMM-estimate. After the contamination is larger than some value ε^* the maximum bias of the MM-estimate is constantly equal to the value of the estimated parameter. This means that the asymptotic value of the estimate becomes 0 independently of the true value of the parameter. This value ε^* correspond to the breakdown point notion proposed by Genton and Lucas [11]. For the AR(1) model the value ε^* depends on ϕ. For the MA(1) model $\varepsilon^* = 0$. Instead , the behavior of the BMM-estimate is different and apparently the estimate does not break down. A surprising feature of its maximum bias curve is that for very large ε the maximum bias starts decreasing. This can be explained as follows: when ε is large, the probability of obtaining a patch of two or more outliers increases. The effect of a patch of outliers is to increase the correlation of the series and therefore, in the case of the AR(1) model with ϕ positive and MA(1) with θ negative it prevents that the parameter further approximates to zero for outliers with fixed size k.We also computed the maximum biases curves for other values of parameters ϕ and θ and the results were similar.

We conjecture that the robust behavior of the BMM-estimate under additive outlier contamination also holds when we observed any contaminated process z_t^ε as given in (3). The reason is that since the BIP-ARMA model includes a built-in filtering to compute the residuals, a small fraction of outliers will affect only a small fraction of residuals. Therefore, since the loss function of the BMM-estimate is bounded, the estimate will not be largely affected by a small fraction of large residuals. We compute maximum bias curves for the case

Estimate	$AR(1)$		$MA(1)$	
	μ	$\phi = 0.5$	μ	$\theta = -0.5$
MLE	0.017	0.0036	0.010	0.0042
MM	0.018	0.0045	0.012	0.0046
BMM	0.018	0.0042	0.012	0.0052
$\mathrm{CTC_B}$	0.017	0.0036	0.011	0.0042
$\mathrm{CTC_3}$	0.017	0.0036	0.011	0.0044
FTAU	0.019	0.0054	0.012	0.0065

Table 1: MSE of the AR(1) and MA(1) models without outliers

of replacement outliers ($w_t^k = k$) obtaining similar results that for the case of additive outliers.

7 A Monte Carlo Study

We have performed a Monte Carlo study to compare several estimates for ARMA models. We have simulated three Gaussian stationary ARMA models considering the case that the series do not contain outliers and the case that the series have 10% of equally spaced in time additive outliers. The sample size in the simulations is 200 and the Monte Carlo study was done with 500 replications.

The estimates considered in this study are (i) the normal conditional maximum likelihood estimate (MLE), (ii) an MM-estimate where the residuals are computed as in a regular ARMA model (MM), (iii) an MM-estimate where the residuals are compared with the ones of a BIP-ARMA model (BMM), (iv) an estimate based on the diagnostic procedure proposed by Chang, Tiao and Chen [5] and which is further described in Chen and Liu [6]. The cutoff point for outlier rejection is chosen by the Bonferroni inequality as $c = \Phi^{-1}(1-(0.05/n))$, where Φ is the $N(0,1)$ distribution function. We denote this estimate by (CTC_B). (v) The same as in (iv) but the cutoff point is $c = 3$ (CTC_3). (vi) The tau filtered estimate proposed by Bianco et al. [1]. We denote this estimate by (FTAU).

The estimates MM and BMM are based on the functions ρ_1 and ρ_2 and η described in Section 5

In Table 1 we show the MSE for the six estimates when the observations come from an AR(1) and a MA(1) model without outliers. Table 2 show the MSE of the same estimates for an ARMA(1,1) model without outliers. The relative efficiency with respect to the MLE varies from 80% to 90% for the estimate BMM in the case of ϕ and θ, from 80 % to 94% for the estimate MM , is practically 100% for the CTC estimates and varies from 65% to 84% for the FTAU. The efficiency of all the estimates of μ is very high.

In Tables 3 and 4 we show the Mean Square Error of estimation (MSE) of the six estimates and the three models with 10% of additive outliers of size

Estimate	μ	$\phi = 0.5$	$\theta = -0.5$
MLE	0.045	0.0062	0.0061
MM	0.050	0.0073	0.0075
BMM	0.051	0.0069	0.0075
$\mathrm{CTC_B}$	0.045	0.0062	0.0061
$\mathrm{CTC_3}$	0.047	0.0061	0.0064
FTAU	0.054	0.0074	0.0082

Table 2: MSE for the ARMA(1,1) models without outliers

	$AR(1)$		$MA(1)$	
Estimate	μ	$\phi = 0.5$	μ	$\theta = -0.5$
MLE	0.189	0.103	0.178	0.128
MM	0.024	0.085	0.0152	0.115
BMM	0.021	0.014	0.0156	0.025
$\mathrm{CTC_B}$	0.185	0.103	0.174	0.130
$\mathrm{CTC_3}$	0.148	0.096	0.136	0.125
FTAU	0.032	0.011	0.020	0.031

Table 3: MSE of the AR(1) and MA(1) models with 10 percent of equally spaced additive outliers of size 4

4 and in Tables 5 and 6 we show the MSE of the six estimates and the three models with 10% of additive outliers of size 6. We observe that for all models the estimate BMM of ϕ and θ behaves much better than those corresponding to the estimates MM, $\mathrm{CTC_B}$ and $\mathrm{CTC_3}$. The performance of the estimates FTAU and BMM are comparable.

The errors of the MSEs shown on these tables are smaller than 15% with probability 0.95. However since the all the estimate were computed with the same samples, the errors of the differences between the MSE of any two estimates are much smaller making comparisons possible .

Estimate	μ	$\phi = 0.5$	$\theta = -0.5$
MLE	0.204	0.023	0.299
MM	0.093	0.021	0.3835
BMM	0.088	0.017	0.060
$\mathrm{CTC_B}$	0.203	0.023	0.300
$\mathrm{CTC_3}$	0.183	0.022	0.313
FTAU	0.082	0.018	0.040

Table 4: MSE of the ARMA(1,1) models with 10 percent of equally spaced additive outliers of size 4

	$AR(1)$		$MA(1)$	
Estimate	μ	$\phi = 0.5$	μ	$\theta = -0.5$
MLE	0.394	0.189	0.380	0.215
MM	0.028	0.132	0.016	0.159
BMM	0.019	0.0048	0.012	0.0065
CTC_B	0.364	0.189	0.345	0.218
CTC_3	0.057	0.047	0.042	0.064
FTAU	0.028	0.0076	0.017	0.021

Table 5: MSE of the AR(1) and MA(1) models with 10 percent of equally spaced additive outliers of size 6

Estimate	μ	$\phi = 0.5$	$\theta = -0.5$
MLE	0.402	0.070	0.374
MM	0.257	0.034	0.585
BMM	0.065	0.011	0.012
CTC_B	0.393	0.069	0.378
CTC_3	0.182	0.042	0.334
FTAU	0.093	0.013	0.025

Table 6: MSE of the ARMA(1,1) models with 10 percent of equally spaced additive outliers of size 6

8 An example

This example deals with a monthly series of inward movement of residential telephone extensions in a fixed geographic area from January 1966 to May 1973 (RESEX). The series was analyzed by Brubacher [3] and by Martin, Samarov and Vandaele [20] , who identified an AR(2) model for the differenced series $y_t = x_t - x_{t-12}$, where x_t is the observed series.

Table 7 displays the value of the estimates MLE, MM, BMM, CTC_3 and the FTAU together with the MAD-scale of the residuals. We can see that the estimated values of the parameters of the MLE and the CTC_3 are quite different from the robust estimates MM, BMM and FTAU. The estimate CTC_B gives the same result as CTC_3 (it detects the same outliers) and it is omitted from the table.

Figure 4 shows the data y_t obtained differentiating the observed data as $y_t = x_t - x_{t-12}$ and the cleaned values as in (10) , which are seen to be almost coincident except at outlier locations.

Estimates	μ	ϕ_1	ϕ_2	MAD
MLE	2.69	0.48	-0.17	1.70
MM	1.18	0.34	0.31	1.43
BMM	1.74	0.42	0.36	1.24
CTC_3	3.44	1.14	-0.74	1.86
FTAU	1.71	0.27	0.49	1.10

Table 7: Estimates of the parameters of the RESX series

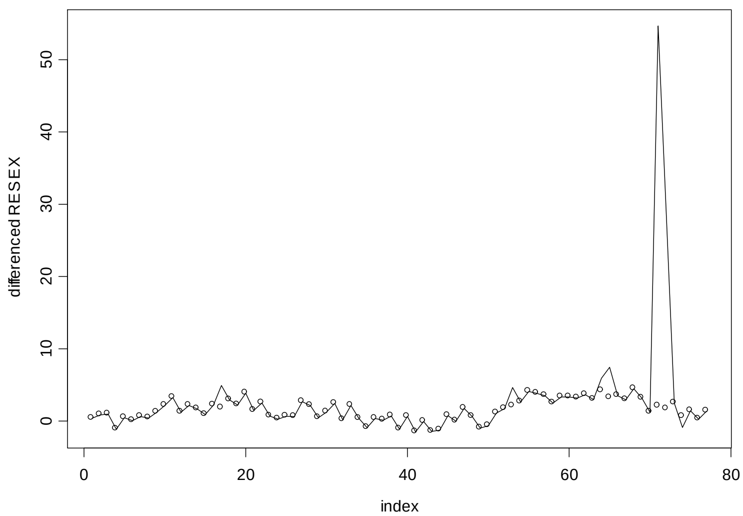

Figure 4: Differenced RESEX Series: Observed (solid line) and Filtered (circles) values

9 Concluding remarks

We have presented two families of estimates for ARMA models: MM-estimates and BMM-estimates. The BMM-estimates uses a mechanism that avoids the propagation of the full effect of the outliers to the subsequent residual innovations. To make this mechanism compatible with consistency when the true model is ARMA, we consider two estimates: one is obtained fitting a regular ARMA model and the other fitting a BIP-ARMA model, where the propagation of the effect of outliers is bounded. Then, the estimate which fits better to the data is selected. We have shown in Sections 6 and 7 that, at least for additive outliers, the BMM-estimates are much more robust than the MM-estimates and quite comparable with the FTAU-estimates. The main advantage of the BMM-estimates over the FTAU-estimates is that an asymptotic theory is now available and this makes inference with BMM-estimates possible. The Monte Carlo results of Section 7 also show that the BMM-estimate compares favorably with the estimate based on the Chen and Liu [6] diagnostic procedure.

10 Appendix

Suppose that we have the infinite sequence of observations

$$\mathbf{Y}_t = (..., y_{t-k}, ..., y_{t-1}, y_t)$$

generated by a stationary and invertible ARMA(p, q) process up to time t with parameter $\boldsymbol{\beta}_0$. Given any $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$ such that the polynomials $\phi(B)$ and $\theta(B)$ have all the roots outside the unit circle, let us define

$$a_t^e(\boldsymbol{\beta}) = \theta^{-1}(B)\phi(B)(y_t - \mu). \tag{28}$$

Then $a_t^e(\boldsymbol{\beta}_0) = a_t$ and $a_t^e(\boldsymbol{\beta})$'s satisfy the following recursive relationship

$$a_t^e(\boldsymbol{\beta}) = y_t - \mu - \sum_{i=1}^{p} \phi_i(y_{t-i} - \mu) + \sum_{i=1}^{q} \theta_i a_{t-i}^e(\boldsymbol{\beta}).$$

In the case that a_t has finite first moment, we have that $a_t^e(\boldsymbol{\beta}) = y_t - E(y_t|\mathbf{Y}_{t-1})$, where the conditional expectation is taken assuming that the true value of the parameter vector is $\boldsymbol{\beta}$.

It is straightforward to derive the following formulas for the for the first and second derivatives of $a_t^e(\boldsymbol{\beta})$

$$\frac{\partial a_t^e(\boldsymbol{\beta})}{\partial \phi_i} = -\theta^{-1}(B)(y_{t-i} - \mu),\ 1 \leq i \leq p, \tag{29}$$

$$\begin{aligned} \frac{\partial a_t^e(\boldsymbol{\beta})}{\partial \theta_j} &= \theta^{-1}(B)a_{t-j}^e(\boldsymbol{\beta}) \\ &= \theta^{-2}(B)\phi(B)(y_{t-j} - \mu),\ 1 \leq j \leq q, \end{aligned} \tag{30}$$

$$\frac{\partial a_t^e(\boldsymbol{\beta})}{\partial \mu} = -\frac{1 - \sum_{i=1}^{p} \phi_i}{1 - \sum_{i=1}^{p} \theta_i}, \tag{31}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial \phi_i \partial \phi_j} = 0, 1 \leq i \leq p, 1 \leq j \leq p, \tag{32}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial \phi_i \partial \theta_j} = -\theta^{-2}(B)(y_{t-j-i} - \mu),\ 1 \leq i \leq p, 1 \leq j \leq q, \tag{33}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial\theta_i\partial\theta_j} = 2\theta^{-3}(B)\phi(B)(y_{t-i-j}-\mu)\ ,\ 1\leq i\leq q, 1\leq j\leq q, \tag{34}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial\phi_i\partial\mu} = \frac{1}{1-\sum_{i=1}^q \theta_i}, 1\leq i\leq p, \tag{35}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial\theta_i\partial\mu} = -\frac{1-\sum_{i=1}^p \phi_i}{(1-\sum_{i=1}^p \theta_i)^2}, 1\leq i\leq q, \tag{36}$$

$$\frac{\partial^2 a_t^e(\boldsymbol{\beta})}{\partial\mu^2} = 0. \tag{37}$$

We will use the following notation. Given a function $g(\mathbf{u}): R^k \rightarrow R$,we define $\nabla g(\mathbf{u})$ as the column vector of dimension k whose i-th element is $\nabla_i g(\mathbf{u}) = \partial g(\mathbf{u})/\partial u_i$ and $\nabla^2 g(\mathbf{u})$ is the $k\times k$ matrix whose (i,j) element is $\nabla_{ij}^2 g(\mathbf{u}) = \partial^2 g(\mathbf{u})/\partial u_i \partial u_j$.

We start proving the following Lemma.

Lemma 1 *Assume y_t satisfies P2. Then for any $d>0$ we have:*

(i) There exists a stationary process W_{0t} such that $\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} |a_t^e(\boldsymbol{\beta})| \leq W_{0t}$ and if $E(|y_t|^2)<\infty$, then $E(|W_{0t}|^2)<\infty$.

(ii) There exists a stationary process W_{1t} such that $\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \|\nabla a_t^e(\boldsymbol{\beta})\| \leq W_{1t}$ and if $E(|y_t|^2)<\infty$, then $E(|W_{1t}|^2)<\infty$.

(iii) There exists a stationary process W_{2t} such that $\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \left\|\nabla^2 a_t^e(\boldsymbol{\beta})\right\| \leq W_{2t}$,where $||A||$ denotes the l_2 norm of matrix A.If $E(|y_t|^2)<\infty$, then $E(|W_{2t}|^2)<\infty$.

Proof.

Since $\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]$, using (28), it is easy to show that there are positive constants k_0, k_1 and $0<\xi<1$,such that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} |a_t^e(\boldsymbol{\beta})| \leq k_0 + k_1\sum_{i=0}^{\infty}\xi^i |y_{t-i}|.$$

Define $W_{0t} = k_0 + k_1\sum_{i=0}^{\infty}\xi^i |y_{t-i}|$. Then by Lemma 1 of Yohai and Maronna [27], W_{0t} is finite.

To prove that if $E(|y_t|^2)<\infty$ then $E(W_{0t}^2)<\infty$, it is enough to show that

$$E\left(\sum_{i=0}^{\infty}\xi^i |y_{t-i}|\right)^2 < \infty,$$

and this follows from

$$\begin{aligned}
E\left(\sum_{i=0}^{\infty}\xi^{i}\left|y_{t-i}\right|\right)^{2} &\leq \sum_{i_2=0}^{\infty}\sum_{i_1=0}^{\infty} \xi^{i_1+i_2}E(|y_{t-i_1}y_{t-i_2}|) \\
&\leq \sum_{i_2=0}^{\infty}\sum_{i_1=0}^{\infty} \xi^{i_1+i_2}E^{1/2}(|y_{t-i_1}|^2)E^{1/2}(|y_{t-i_2}|^2) \\
&= E(|y_t|^2)\sum_{i_2=0}^{\infty}\sum_{i_1=0}^{\infty} \xi^{i_1+i_2} = E(|y_t|^2)\left(\sum_{i=0}^{\infty}\xi^{i}\right)^{2}.
\end{aligned}$$

Therefore (i) follows.

From (29)-(31) and (32)-(37) we can prove (ii) and (iii) respectively using the same arguments as in the proof of (i).

In the next Lemma we prove the Fisher Consistency of the S-estimate when we have all the past observations.

Lemma 2 *Assume that y_t satisfies condition P2 with innovations satisfying P3. Assume that ρ_1 is a bounded function satisfying condition P1, define $s(\boldsymbol{\beta})$ by*

$$E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s(\boldsymbol{\beta})}\right)\right) = b. \tag{38}$$

Then if $\boldsymbol{\beta}\in\mathcal{B}$ and $\boldsymbol{\beta}\neq\boldsymbol{\beta}_0$ we have $s_0 = s(\boldsymbol{\beta}_0) < s(\boldsymbol{\beta})$.

Proof.

Let $\boldsymbol{\beta} = (\boldsymbol{\phi},\boldsymbol{\theta},\mu) \neq \boldsymbol{\beta}_0 = (\boldsymbol{\phi}_0,\boldsymbol{\theta}_0,\mu_0)$.We have

$$\begin{aligned}
a_t^e(\boldsymbol{\beta}) &= \theta^{-1}(B)\phi(B)(y_t-\mu) \\
&= \theta^{-1}(B)\phi(B)(y_t-\mu_0) + \theta^{-1}(B)\phi(B)(\mu_0-\mu) \\
&= \omega(B)a_t + c(\mu_0-\mu), \qquad (39)
\end{aligned}$$

where

$$\begin{aligned}
\omega(B) &= \theta^{-1}(B)\theta_0(B)\phi_0^{-1}(B)\phi(B) \\
&= 1+\sum_{i=1}^{\infty}\omega_i B^i
\end{aligned}$$

and

$$c = \frac{1-\sum_{i=1}^{p}\phi_i}{1-\sum_{i=1}^{q}\theta_i} \neq 0.$$

Put

$$\Delta_t(\boldsymbol{\beta}) = \sum_{i=1}^{\infty}\omega_i a_{t-i} + c(\mu_0-\mu). \tag{40}$$

Then, from (39) we obtain

$$E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right) = E\left(\rho_1\left(\frac{1}{s_0}\left(a_t + \Delta_t(\boldsymbol{\beta})\right)\right)\right).$$

Let $S(p,q)$

$$S(p,q) = E\left(\rho_1\left(\frac{a_t + p}{q}\right)\right),$$

and observe that $S(p,q)$ is decreasing in q. Lemma 3.1 of Yohai [25], showed that if P1 and P3 hold then for all $p \neq 0$ and $q \neq 0$

$$S(0,q) < S(p,q). \tag{41}$$

Then, since a_t and $\Delta_t(\boldsymbol{\beta})$ are independent we have

$$\begin{aligned} E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right) &= E(S(\Delta_t(\boldsymbol{\beta}), s_0)) \\ &\geq E(S(0, s_0)) \\ &= E\left(\rho_1\left(\frac{a_t}{s_0}\right)\right) = b \end{aligned}$$

and the equality holds if and only if $\Delta_t(\boldsymbol{\beta}) = 0$ a.s.. Because of the identifiability of the ARMA model, this occurs if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}^0$. Then $\boldsymbol{\beta} \neq \boldsymbol{\beta}^0$ implies

$$E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right) = E(S(\Delta_t(\boldsymbol{\beta}), s_0)) > b$$

and therefore we have $s(\boldsymbol{\beta}) > s_0 = s(\boldsymbol{\beta}_0)$.

The next two Lemmas are very well known properties of difference equation. They are proved for sake of completeness.

Lemma 3. *Consider the difference equation*

$$z_t = \sum_{i=1}^{k} \lambda_i z_{t-i} + d \tag{42}$$

and assume that $\sum_{i=1}^{k} \lambda_i \neq 1$. *Let* z_t, $t \geq t_0 + 1$ *be a solution of (42) corresponding to given initial values* $z_{t_0}, z_{t_{0-1}}, \cdots z_{t_0-k}$. *Then it holds*

$$|z_t - z^*| \leq \left\| A(\boldsymbol{\lambda})^{t-t_0} \right\| \left(|z_{t_0} - z^*|^2 + ... + |z_{t_0-k+1} - z^*|^2\right)^{1/2}, \tag{43}$$

where $z^* = d/(1 - \sum_{i=1}^{k} \lambda_i)$,

$$A(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & ... & \lambda_{k-1} & \lambda_k \\ 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 1 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & 1 & 0 \end{pmatrix} \tag{44}$$

and $||A||$ denotes the l_2 norm of the matrix A.

Proof.

It is enough to show the Lemma for $t = t_0 + 1$. Call

$$\mathbf{z}_t = (z_t, z_{t-1}, ...z_{t-k+1})', \mathbf{z}^* = (z^*, ..., z^*)', \mathbf{d} = (d, 0...0)'.$$

Then it is easy to check that

$$\mathbf{z}_t = A(\boldsymbol{\lambda})\mathbf{z}_{t-1} + \mathbf{d},\ \mathbf{z}^* = A(\boldsymbol{\lambda})\mathbf{z}^* + \mathbf{d}$$

and then

$$\mathbf{z}_t - \mathbf{z}^* = A(\boldsymbol{\lambda})(\mathbf{z}_{t-1} - \mathbf{z}^*).$$

Therefore we have that

$$||\mathbf{z}_t - \mathbf{z}^*|| \leq \|A(\boldsymbol{\lambda})\| \, ||\mathbf{z}_{t-1} - \mathbf{z}^*||$$

and then (43) holds for $t = t_0 + 1$. This proves the Lemma.

Given $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_k) \in R^k$, let $\lambda(x)$ be the polynomial

$$\lambda(x) = 1 - \sum_{i=1}^{k} \lambda_i x^i.$$

Lemma 4 *Given $\varepsilon > 0$, let C_ε be the set of all $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_k)$ such that the polynomial $\lambda(B) = 1 - \lambda_1 B - ... - \lambda_k B^K$ has all its roots have modulus larger or equal than $1 + \varepsilon$ and let $A(\boldsymbol{\lambda})$ be the matrix as defined in (44). Then, there exists $0 < \nu < 1$ and a positive constant C such that*

$$\sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| A(\boldsymbol{\lambda})^t \right\| \leq C\nu^t. \tag{45}$$

Proof.

Using the Jordan canonical form we can write

$$A(\boldsymbol{\lambda}) = B^{-1}(\boldsymbol{\lambda})(D(\boldsymbol{\lambda}) + N(\boldsymbol{\lambda}))B(\boldsymbol{\lambda}),$$

where $D(\boldsymbol{\lambda})$ is a diagonal matrix that has the eigenvalues of $A(\boldsymbol{\lambda})$ in the diagonal, and $N(\boldsymbol{\lambda})$ is a nilpotent matrix of $0'$s and $1'$s satisfying that $N(\boldsymbol{\lambda})^k$ is the null matrix. and such that $||N(\boldsymbol{\lambda})|| \leq 1$.Moreover$N(\boldsymbol{\lambda})$ and $D(\lambda)$ commute. Put $C(\lambda) = D(\boldsymbol{\lambda}) + N(\boldsymbol{\lambda})$, then

$$A(\lambda)^t = B^{-t}(\lambda)C^t(\lambda)B^t(\lambda). \tag{46}$$

Since $N(\boldsymbol{\lambda})$ and $D(\lambda)$ commute and $N(\boldsymbol{\lambda})^{\ k}$ is the null matrix, we get

$$\sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| C^t(\boldsymbol{\lambda}) \right\| \leq \sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \sum_{j=0}^{k-1} \binom{t}{j} \|N(\boldsymbol{\lambda})\|^j \|D(\boldsymbol{\lambda})\|^{t-j} \tag{47}$$

Let $\zeta_i(\boldsymbol{\lambda}), 1 \leq i \leq k$, be the roots of $\boldsymbol{\lambda}(B)$.Since the eigenvalues of $A(\lambda)$ are $1/\zeta_i(\boldsymbol{\lambda}), 1 \leq i \leq k$, from (47) we obtain

$$\begin{aligned}
\sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| C^t(\boldsymbol{\lambda}) \right\| &\leq \sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \sum_{j=0}^{k-1} \binom{t}{j} \left(\frac{1}{1+\varepsilon} \right)^{t-j} && (48) \\
&\leq \left(\frac{1}{1+\varepsilon} \right)^{t-k+1} \\
&\leq \left(\frac{1}{1+\varepsilon} \right)^{-k+1} \left(\frac{1}{1+\varepsilon} \right)^{t}. && (49)
\end{aligned}$$

For $0 \leq t \leq k-1$, we can write

$$\begin{aligned}
\sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| C^t(\boldsymbol{\lambda}) \right\| &\leq \sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \sum_{j=0}^{t} \binom{t}{j} \left\| N(\boldsymbol{\lambda}) \right\|^j \left\| D(\boldsymbol{\lambda}) \right\|^{t-j} \\
&\leq \sum_{j=0}^{t} \binom{t}{j} \frac{1}{(1+\varepsilon)^{t-j}} \leq 2^{t-1}. && (50)
\end{aligned}$$

Then from (49) and (50) it is easy to prove that there exists C_1 such that

$$\sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| C^t(\boldsymbol{\lambda}) \right\| \leq C_1 \nu^t, . \tag{51}$$

where $\nu = 1/(1+\varepsilon)$. Put $C_2 = \sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| B(\lambda) \right\|$ and $C_3 = \sup_{\boldsymbol{\lambda} \in \mathcal{C}_\varepsilon} \left\| B^{-1}(\lambda) \right\|$ and $C = C_1 C_2 C_3$. Then, from (46) and (51) we get (45).

Lemma 5 *Under the assumptions of Theorem 1, for any $d > 0$ we have,*

$$\lim_{n \to \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d,d]} \left| S_n(\mathbf{a}_n(\boldsymbol{\beta})\ ) - s(\boldsymbol{\beta}) \right| = 0 \text{ a.s..}$$

Proof.

It is easy to show that $s(\boldsymbol{\beta})$ is continuous and positive. Let

$$h_1 = \inf_{\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d,d]} s(\boldsymbol{\beta})\ ,\ h_2 = \sup_{\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d,d]} s(\boldsymbol{\beta}).$$

Then $h_1 > 0$ and $h_2 < \infty$. From Lemma 2 of Muler and Yohai [22] we have that

$$\begin{aligned}
&\lim_{n \to \infty} \sup_{\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d,d]\ , c \in [h_1/2, 2h_2]} \left| \sum_{t=p+1}^{n} \frac{\rho_1\left(a_t^e(\boldsymbol{\beta})/c\right)}{n-p} - E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{c} \right) \right) \right| \\
&= \quad 0 \text{ a.s..} && (52)
\end{aligned}$$

Since ψ_1 is bounded and continuous, there exists $C > 0$ such that for $t \geq p+1$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]\ ,c\in[h_1/2,2h_2]} \sum_{t=p+1}^{n}\left|\rho_1\left(\frac{a_t(\boldsymbol{\beta})}{c}\right)-\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{c}\right)\right|$$

$$\leq \quad C\sum_{t=p+1}^{n}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left|a_t(\boldsymbol{\beta})-a_t^e(\boldsymbol{\beta})\right|. \tag{53}$$

Put $g_t(\boldsymbol{\beta}) = a_t^e(\boldsymbol{\beta}) - a_t(\boldsymbol{\beta})$, it is easy to verify that for $t \geq p+1$,

$$g_t(\boldsymbol{\beta}) = \sum_{i=1}^{q}\theta_i g_{t-i}(\boldsymbol{\beta})$$

with $g_{q+1-i}(\boldsymbol{\beta}) = a_{q+1-i}^e(\boldsymbol{\beta}), 1 \leq i \leq q$. Then by the definition of $\mathcal{B}_0$ and Lemmas 3 and 4 there exists $0 < \nu < 1$ and a positive constant C_1 such that for $t \geq p+1$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}|g_t(\boldsymbol{\beta})| \leq C_1\nu^t \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left(\sum_{i=1}^{q}a_{q+1-i}^{e2}(\boldsymbol{\beta})\right)^{1/2},$$

and by Lemma 1 (i) we get for $t \geq p+1$,

$$\begin{aligned}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}|a_t(\boldsymbol{\beta})-a_t^e(\boldsymbol{\beta})| \quad &\leq \quad C_1\nu^t\left(\sum_{i=1}^{q}W_{0,q+1-i}^2\right)^{1/2}\\ &\leq \quad \nu^t Z,\end{aligned} \tag{54}$$

where Z is the random variable $C_1\left(\sum_{i=1}^{q}W_{0,q+1-i}^2\right)^{1/2}$.

Therefore from (52), (53) and (54) we have

$$\lim_{n\to\infty}\ \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]\ ,c\in[h_1/2,2h_2]}\left|\sum_{t=p+1}^{n}\frac{\rho_1\left(a_t(\boldsymbol{\beta})/c\right)}{n-p}-E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{c}\right)\right)\right| = 0 \text{ a.s..} \tag{55}$$

Let $0 \leq \varepsilon \leq h_1/2$ and define

$$g_1(\boldsymbol{\beta}) = E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s(\boldsymbol{\beta})+\varepsilon}\right)\right),\ g_2(\boldsymbol{\beta}) = E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{s(\boldsymbol{\beta})-\varepsilon}\right)\right).$$

By (38) we have that $g_1(\boldsymbol{\beta}) < b$ and $g_2(\boldsymbol{\beta}) > b$. Since $\mathcal{B}_0$ is a compact set and g_1 and g_2 are continuous, we have

$$\kappa_1 = \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} g_1(\boldsymbol{\beta}) < b, \kappa_2 = \inf_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} g_2(\boldsymbol{\beta}) > b.$$

Let $\delta = \min(b-\kappa_1, \kappa_2 - b)$. From (55) there exists n_0 such that for all $n \geq n_0$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]\ ,c\in[h_1/2,2h_2]} \left| \frac{1}{n-p} \sum_{t=p+1}^{n} \rho_1\left(\frac{a_t(\boldsymbol{\beta})}{c}\right) - E\left(\rho_1\left(\frac{a_t^e(\boldsymbol{\beta})}{c}\right)\right)\right| \leq \frac{\delta}{2}.$$

Therefore, for all $n \geq n_0$ we get,

$$\inf_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \frac{1}{n-p} \sum_{t=p+1}^{n} \rho_1\left(\frac{a_t(\boldsymbol{\beta})}{s(\boldsymbol{\beta})-\varepsilon}\right) \geq \kappa_2 - \frac{\delta}{2} \geq b + \frac{\delta}{2}.$$

Similarly we have,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \frac{1}{n-p} \sum_{t=p+1}^{n} \rho_1\left(\frac{a_t(\boldsymbol{\beta})}{s(\boldsymbol{\beta})+\varepsilon}\right) \leq \kappa_1 + \frac{\delta}{2} \leq b - \frac{\delta}{2},$$

and hence from the monotonicity of $\rho_1(|u|)$ we get,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} |S_n(\mathbf{a}_n(\boldsymbol{\beta})) - s(\boldsymbol{\beta})| \leq \varepsilon.$$

This proves the Lemma.

Lemma 6. *Under the assumptions of Theorem 1, there exists $d > 0$ satisfying*

$$\lim_{n\to\infty} \inf \inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} S_n(\mathbf{a}_n(\boldsymbol{\beta})) > s_0 + 1 \text{ a.s..}$$

Proof.

Given $\boldsymbol{\beta} = (\mu, \boldsymbol{\phi}, \boldsymbol{\theta})$ with $(\boldsymbol{\phi}, \boldsymbol{\theta}) \in\mathcal{B}_0$, let us call $\vartheta_t(\boldsymbol{\beta}) = a_t(\boldsymbol{\beta}) - a_t(\boldsymbol{\phi}, \boldsymbol{\theta}, 0)$. From (13), it is easy to show that $\vartheta_t(\boldsymbol{\beta})$ satisfy for $t \geq p+1$ the difference equation,

$$\vartheta_t(\boldsymbol{\beta}) = \sum_{i=1}^{q} \theta_i \vartheta_{t-i}(\boldsymbol{\beta}) + \mu\left(1 - \sum_{i=1}^{p} \phi_i\right) \tag{56}$$

with initial conditions $\vartheta_{p+1-i}(\boldsymbol{\beta}) = 0$, $1 \leq i \leq q$. Moreover, it is easy to verify that

$$\vartheta_t(\boldsymbol{\beta}) = \mu\vartheta_t(\boldsymbol{\phi}, \boldsymbol{\theta}, 1).$$

Using the definition of $\mathcal{B}_0$, there exists $\delta > 0$ and $K_1 > 0$ such that for all $(\boldsymbol{\phi}, \boldsymbol{\theta}) \in\mathcal{B}_0$,

$$\delta \leq \frac{1 - \sum_{i=1}^{q} \phi_i}{1 - \sum_{i=1}^{q} \theta_i} \leq K_1. \tag{57}$$

Then, by Lemmas 3 and 4 there exists $0 < \nu < 1$ and $K_2 > 0$ such that for $t \geq p+1$,

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \left| \vartheta_t(\boldsymbol{\phi},\boldsymbol{\theta},1) - \frac{1-\sum_{i=1}^{q}\phi_i}{1-\sum_{i=1}^{q}\theta_i} \right| \leq \nu^t K_2,$$

and so there exists t_0 such that for $t \geq t_0$, we have

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \left| \vartheta_t(\boldsymbol{\phi},\boldsymbol{\theta},1) - \frac{1-\sum_{i=1}^{q}\phi_i}{1-\sum_{i=1}^{q}\theta_i} \right| \leq \frac{\delta}{2}. \tag{58}$$

Then, from (57) and (58) we get,

$$\inf_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \left|\vartheta_t\left(\boldsymbol{\beta}\right)\right| \geq \frac{\delta}{2}\left|\mu\right|. \tag{59}$$

From (54) there exists a random variable Z, and $0 < \nu < 1$ such that for all $t \geq p+1$ it holds

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} |a_t^e(\boldsymbol{\phi},\boldsymbol{\theta},0) - a_t(\boldsymbol{\phi},\boldsymbol{\theta},0)| \leq \nu^t Z \tag{60}$$

and by Lemma 1 (i) we obtain

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} |a_t^e(\boldsymbol{\phi},\boldsymbol{\theta},0)| \leq W_{0,t}, \tag{61}$$

where $W_{0,t}$ is a stationary process.

Since $\sup\rho_1 > b$ and $\lim_{x\to\infty}\rho_1(|x|) = \sup\rho_1$, there exist $k_0 > 0$ and $\lambda > 1$ such that for all x satisfying $|x| \geq k_0$ it holds

$$\rho_1(x) \geq \lambda b. \tag{62}$$

Let m be such that

$$P(W_{0,t} < m/2\ ) > \frac{1}{\lambda}, \tag{63}$$

$$k = \max\left(\frac{m}{s_0+1}, k_0\right) \tag{64}$$

and $d \geq 4(s_0+1)k/\delta$. Then using (59) we get for all $t \geq t_0$

$$\inf_{|\mu|>d,(\boldsymbol{\theta},\boldsymbol{\phi})\in\mathcal{B}_0} |\vartheta_t(\boldsymbol{\beta})| \geq 2(s_0+1)k. \tag{65}$$

Since ρ_1 satisfies property P1, it holds

$$\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\frac{a_t(\boldsymbol{\beta})}{s_0+1}\right)$$
$$\geq \frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\left|\frac{a_t(\boldsymbol{\beta})}{s_0+1}\right|\right) I(A_t)I(B_t), \qquad (66)$$

where $A_t = \{W_{0t} < m/2\}$ and $B_t = \{\nu^t Z < m/2\}$.

From (60) and (61) we can write,

$$|a_t(\boldsymbol{\beta})| \geq |\vartheta_t(\boldsymbol{\beta})| - (W_{0,t} + \nu^t Z).$$

Then, from (64) and (65) we obtain for all $t \geq t_0$ that

$$\left\{\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} |a_t(\boldsymbol{\beta})| > k(s_0+1)\right\}$$
$$\supset \left\{W_{0,t} + C\nu^t Z < k(s_0+1)\right\} \supset A_t \cap B_t. \qquad (67)$$

Since $\rho_1 \geq 0$, and $\rho_1(|u|)$ is non decreasing, from (66) and (67), we get

$$\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\left|\frac{a_t(\boldsymbol{\beta})}{s_0+1}\right|\right)$$
$$\geq \frac{1}{n-p}\sum_{t=t_0}^{n}\rho_1(k)\, I(A_t)I(B_t). \qquad (68)$$

With probability 1, there exists $t_1 \geq t_0$ such that $I(B_t) = 1$ for all $t \geq t_1$.Then

$$\begin{aligned}\frac{\rho_1(k)}{n-p}\sum_{t=t_0}^{n} I(A_t)I(B_t) &\geq \frac{\rho_1(k)}{n-p}\sum_{t=t_1+1}^{n} I(A_t)\\ &\geq \frac{\rho_1(k)}{n-p}\sum_{t=p+1}^{n} I(A_t) - \frac{t_1-p}{n-p}\rho_1(k). \qquad (69)\end{aligned}$$

Since W_{0t} is an ergodic and stationary process, from (63) we have

$$\lim_{n\to\infty}\frac{1}{n-p}\sum_{t=p+1}^{n} I(A_t) = P(A_t) > \frac{1}{\lambda}.$$

Then, from (62), (64), (66), (68) and (69) we get

$$\lim\inf_{n\to\infty}\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\frac{a_t(\boldsymbol{\beta})}{s_0+1}\right) \geq \frac{\rho_1(k)}{\lambda} \geq b$$

and the Lemma follows.

Proof of Theorem 1.

Take $\varepsilon > 0$ arbitrarily small and let d be as in Lemma 6. By the dominated convergence theorem it is easy to show that $s(\boldsymbol{\beta})$ is continuous. Then by Lemma 2, there exists $0 < \gamma < 1$ such that

$$\min_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],||\boldsymbol{\beta}-\boldsymbol{\beta}_9||\geq\varepsilon} s(\boldsymbol{\beta}) \geq s_0 + \gamma.$$

Therefore by Lemma 5, there exist n_1 such that for $n \geq n_1$

$$\min_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\geq\varepsilon} S_n(\boldsymbol{\beta}) \geq s_0 + \gamma/2$$

and

$$S_n(\boldsymbol{\beta}_0) \leq s_0 + \gamma/4.$$

Moreover by Lemma 6, there exists n_2 such that for $n \geq n_2$

$$\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} S_n(\mathbf{a}_n(\boldsymbol{\beta}) > s_0 + \gamma \text{ a.s..}$$

Therefore, for $n \geq \max(n_1, n_2)$ it holds that $||\widehat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}_0|| < \varepsilon$ and this proves the Theorem.

The next three Lemmas will be used to prove Theorem 2.

Lemma 7. *Assume that y_t satisfies condition P2. Given $d > 0$ and $\widetilde{\sigma} > 0$, there exist constants $C > 0$ and $0 < \nu < 1$ such that*

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sup_{0<\sigma\leq\widetilde{\sigma}} \left|a_t^b(\boldsymbol{\beta},\sigma) - y_t\right| \leq C\left(\widetilde{\sigma} + \nu^t \sum_{i=1}^{p} |y_i|\right),\ t \geq p+1.$$

Proof. Given $\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d, d]$ and $\sigma \leq \widetilde{\sigma}$, let us define for $t \geq p+1$,

$$\upsilon_t(\boldsymbol{\beta},\sigma) = a_t^b(\boldsymbol{\beta},\sigma) - y_t,\ D_t(\boldsymbol{\beta},\sigma) = \sigma \sum_{i=1}^{r} (\theta_i - \phi_i)\,\eta\left(a_{t-i}^b(\boldsymbol{\beta},\sigma)/\sigma\right) \tag{70}$$

and $\upsilon_t(\boldsymbol{\beta},\sigma) = -y_t$ for $1 \leq t \leq p$. From (12), $\upsilon_t(\boldsymbol{\beta},\sigma)$ satisfy for $t \geq p+1$ the recursive equation,

$$\upsilon_t(\boldsymbol{\beta},\sigma) = \sum_{i=1}^{p} \phi_i \upsilon_{t-i}(\boldsymbol{\beta},\sigma) + D_t(\boldsymbol{\beta},\sigma) - \mu(1 - \sum_{i=1}^{p}\phi_i),\ t \geq p+1. \tag{71}$$

Define

$$\begin{aligned}
\boldsymbol{\upsilon}_t(\boldsymbol{\beta},\sigma) &= (\upsilon_t(\boldsymbol{\beta},\sigma), \upsilon_{t-1}(\boldsymbol{\beta},\sigma), ..., \upsilon_{t-p+1}(\boldsymbol{\beta},\sigma))', \\
\mathbf{D}_t(\boldsymbol{\beta},\sigma) &= (D_t(\boldsymbol{\beta},\sigma) - \mu(1 - \sum_{i=1}^{p}\phi_i), 0...0)'
\end{aligned}$$

and $A(\boldsymbol{\phi})$ as defined in (44). Then,

$$\boldsymbol{\upsilon}_t(\boldsymbol{\beta},\sigma) = A(\boldsymbol{\phi})\boldsymbol{\upsilon}_{t-1}(\boldsymbol{\beta},\sigma) + \mathbf{D}_t(\boldsymbol{\beta},\sigma)$$

and it is easy to show inductively that for $t \geq p+1$,

$$\boldsymbol{\upsilon}_t(\boldsymbol{\beta},\sigma) = A(\boldsymbol{\phi})^{t-p}\boldsymbol{\upsilon}_p + \sum_{i=0}^{t-p-1} A^i(\boldsymbol{\phi})\mathbf{D}_{t-i}(\boldsymbol{\beta},\sigma). \tag{72}$$

Since $\mathcal{B}_0$ is compact , η is bounded, and $|\mu| \leq d$ it follows that there exists a constant D such that for $t \geq p+1$,

$$\begin{aligned} & \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sup_{0<\sigma\leq\widetilde{\sigma}} \|\mathbf{D}_t(\boldsymbol{\beta},\sigma)\| \\ \leq & \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sup_{0<\sigma\leq\widetilde{\sigma}} |D_t(\boldsymbol{\beta},\sigma)| + (1-\sum_{i=1}^{p}\phi_i)\,|\mu| \leq D\widetilde{\sigma} + d. \end{aligned} \tag{73}$$

Then, from (16), (72), Lemmas 3 and 4 we have that there exists positive constants C_1, and $0<\nu<1$ such that for $t \geq p+1$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sup_{0<\sigma\leq\widetilde{\sigma}} \|\boldsymbol{\upsilon}_t(\boldsymbol{\beta},\sigma)\| \leq C_1\nu^t\,\|\boldsymbol{\upsilon}_p\| + \frac{C_1}{1-\nu}\,(D\widetilde{\sigma}+d)\,. \tag{74}$$

This proves the Lemma.

Lemma 8. *Under the assumptions of Theorem 2, given* $d>0$, *there exists* $\delta>0$ *such that*

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} S_n\left(\mathbf{a}_n^b\left(\boldsymbol{\beta},\widehat{\boldsymbol{\sigma}}(\boldsymbol{\phi},\boldsymbol{\theta})\right)\right) > s_0+\delta \text{ a.s.}.$$

Proof. Since by (19), $\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}) \leq \widehat{\sigma}_y$, using Lemma 7 we can find a constant $C>0$ and $0<\nu<1$ such that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \left|a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})) - y_t\right| \leq C\left(\widehat{\sigma}_y + \nu^t\sum_{i=1}^{p}|y_i|\right).$$

Since $\lim_{n\to\infty}\widehat{\sigma}_y = \sigma_y$ a.s., with probability one there exists t_0 large enough such that for all $t \geq t_0$,

$$\widehat{\sigma}_y + \nu^t\sum_{i=1}^{p}|y_i| \leq 2\sigma_y \text{ a.s.}. \tag{75}$$

Then, calling $D = 2C\sigma_y$ we have for all $t \geq t_0$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \left|a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})) - y_t\right| \leq D. \tag{76}$$

We can write $y_t = \mu_0 + a_t + v_{t-1}$, where v_{t-1} is an stationary process that depends on $a_{t'}$, $t' < t$. Since y_t is not white noise and the distribution of a_t is unbounded, we have that v_t has an unbounded distribution too. Put

$$u_{t-1}(\boldsymbol{\beta}, \sigma) = \mu_0 + v_{t-1} + (a_t^b(\boldsymbol{\beta},\sigma) - y_t). \tag{77}$$

Then, from (12), $u_{t-1}(\boldsymbol{\beta}, \sigma)$ also depends on $a_{t'}$, $t' < t$. We can write

$$a_t^b(\boldsymbol{\beta},\sigma) = y_t + (a_t^b(\boldsymbol{\beta},\sigma) - y_t) = a_t + u_{t-1}(\boldsymbol{\beta}, \sigma), \tag{78}$$

and observe that (76) and (77) imply that for $t \geq t_0$ we have

$$\left\{ \inf_{\boldsymbol{\beta} \in \mathcal{B}_0 \times [-d,d]} |u_{t-1}(\boldsymbol{\beta}, \sigma)| \geq 1 \right\} \supset \{|v_{t-1}| > D + |\mu_0| + 1\}. \tag{79}$$

Since v_t has an unbounded distribution, we have that

$$\gamma = P(|v_t| \geq D + |\mu_0| + 1) > 0.$$

According to definition of s_0 in Theorem 1, we have that $E\left(\rho_1\left(a_t/s_0\right)\right) = b$ and in Lemma 3.1 of Yohai [25] it is shown that

$$E\left(\rho_1\left((a_t + u)/s_0\right)\right) > b$$

for all $u \neq 0$. This implies that

$$\inf_{|u| \geq 1} E\left(\rho_1\left((a_t + u)/s_0\right)\right) > b.$$

Then, since

$$(1-\gamma) E\left(\rho_1\left(\frac{a_t}{s_0}\right)\right) + \gamma \inf_{|u| \geq 1} E\left(\rho_1\left(\frac{a_t + u}{s_0}\right)\right) > b,$$

we can find $\delta > 0$ such that

$$(1-\gamma) E\left(\rho_1\left(\frac{a_t}{s_0 + \delta}\right)\right) + \gamma \inf_{|u| \geq 1} E\left(\rho_1\left(\frac{a_t + u}{s_0 + \delta}\right)\right) \geq b + \delta. \tag{80}$$

Put

$$h(u) = E\left(\rho_1\left(\frac{a_t + u}{s_0 + \delta}\right)\right)$$

and define

$$\begin{aligned} R_t(\boldsymbol{\beta}, \sigma) &= \rho_1\left(\frac{a_t^b(\boldsymbol{\beta},\sigma)}{s_0 + \delta}\right) - h(u_{t-1}(\boldsymbol{\beta}, \sigma)) \\ &= \rho_1\left(\frac{a_t + u_{t-1}(\boldsymbol{\beta}, \sigma)}{s_0 + \delta}\right) - h(u_{t-1}(\boldsymbol{\beta}, \sigma)) \end{aligned}$$

It is easy to verify that $R_t(\beta,\sigma)$ is a bounded martingale difference sequence. Then, by the law of large numbers for martingale differences, see for instance Theorem 20.10 of Davidson [7], we get that

$$\frac{1}{n-p}\sum_{t=p+1}^{n} R_t(\boldsymbol{\beta},\sigma) = 0 \text{ a.s..} \tag{81}$$

Using a compactness argument for all $\varepsilon > 0$ we can find $(\beta_i, \sigma_i, \delta_i)$, $1 \leq i \leq m_0$,with $\beta_i \in \mathcal{B}_0 \times [-d, d]$, $\sigma_i \leq 2\sigma_y$, such that if we define

$$V_i = \{(\boldsymbol{\beta},\sigma) : ||\boldsymbol{\beta} - \boldsymbol{\beta}_i|| + |\sigma - \sigma_i| \leq \delta_i\},$$

we have that $\cup_{i=1}^{m_0} V_i \supset \mathcal{B}_0 \times [-d,d] \times [0, 2\sigma_y]$ and

$$\sup_{(\boldsymbol{\beta},\boldsymbol{\sigma})\in V_i} \left| \frac{1}{n-p}\sum_{t=p+1}^{n} (R_t(\boldsymbol{\beta},\sigma) - R_t(\boldsymbol{\beta}_i,\sigma_i)) \right| \leq \varepsilon$$

This last inequality and (81) imply that

$$\lim_{n\to\infty}\sup \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\ \sigma\leq 2\sigma_y} \left| \frac{1}{n-p}\sum_{t=p+1}^{n} R_t(\boldsymbol{\beta},\sigma) \right| \leq \varepsilon \text{ a.s.,}$$

and since this hold for all $\varepsilon > 0$, we get

$$\lim_{n\to\infty}\sup \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\ \sigma\leq 2\sigma_y} \left| \frac{1}{n-p}\sum_{t=p+1}^{n} R_t(\boldsymbol{\beta},\sigma) \right| = 0 \text{ a.s..} \tag{82}$$

Put

$$\gamma_n = \frac{1}{n-p}\sum_{t=p+1}^{n} I(|v_{t-1}| \geq D + |\mu_0| + 1).$$

Then, we get

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\sigma\leq 2\sigma_y} \frac{1}{n-p}\sum_{t=p+1}^{n} h(u_{t-1}(\boldsymbol{\beta},\sigma)) \geq (1-\gamma_n)h(0) + \gamma_n \inf_{|u|\geq 1} h(u),$$

and therefore since $\gamma_n \to \gamma$,a.s. by (80) we have

$$\begin{aligned} \lim_{n\to\infty}\inf \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\sigma\leq 2\sigma_y} \frac{1}{n-p}\sum_{t=p+1}^{n} h(u_{t-1}(\boldsymbol{\beta},\sigma)) &\geq (1-\gamma)h(0) + \gamma \inf_{|u|\geq 1} h(u) \\ &\geq b + \delta \text{ a.s.} \end{aligned}$$

and then from (82) we have

$$\lim_{n\to\infty}\inf \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\sigma\leq 2\sigma_y} \sum_{t=p+1}^{n} \rho_1\left(\frac{a_t^b(\boldsymbol{\beta},\sigma)}{s_0+\delta}\right) \geq b + \delta.$$

Then, since $\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}) \leq \widehat{\sigma}_y$ by (75), the Lemma follows.

Lemma 9. *Under the assumptions of Theorem 2, there exists $d > 0$ such that,*

$$\liminf_{n\to\infty} \inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} S_n(\mathbf{a}_n^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))\ ) > s_0 + 1 \text{ a.s..}$$

Proof.

Denote by $\mathbf{A}_i^k(\boldsymbol{\phi})$ and $A_{i,j}^k(\boldsymbol{\phi})$ the i-th raw and (i,j) element of the matrix $A^k(\boldsymbol{\phi})$ respectively, where $A(\boldsymbol{\phi})$ is defined in (44). Then, according to (72) we have

$$\begin{aligned} & a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})) - y_t \\ = \quad & \mathbf{A}_1^{t-p}(\boldsymbol{\phi})\boldsymbol{\upsilon}_p + \sum_{i=0}^{t-p-1} A_{1,1}^i(\boldsymbol{\phi}) \left(D_{t-i}(\boldsymbol{\beta},\widehat{\boldsymbol{\sigma}}(\boldsymbol{\phi},\boldsymbol{\theta})) - \mu(1-\sum_{i=1}^{p}\phi_i) \right), \end{aligned}$$

where $\boldsymbol{\upsilon}_p$ and $D_t(\boldsymbol{\beta},\widehat{\boldsymbol{\sigma}}(\boldsymbol{\phi},\boldsymbol{\theta}))$ are defined in (70). Then,

$$\begin{aligned} & \left|a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))\right| \\ \geq \quad & -\left|\mathbf{A}_1^{t-p}(\boldsymbol{\phi})\boldsymbol{\upsilon}_p\right| - \left|\sum_{i=0}^{t-p-1} A_{1,1}^i(\boldsymbol{\phi}) D_{t-i}(\boldsymbol{\beta},\widehat{\boldsymbol{\sigma}}(\boldsymbol{\phi},\boldsymbol{\theta})\right| \\ & + \left|\sum_{i=0}^{t-p-1} A_{1,1}^i(\boldsymbol{\phi})(1-\sum_{i=1}^{p}\phi_i)\mu\right| - |y_t| \\ \geq \quad & -\left\|A^{t-p}(\boldsymbol{\phi})\right\| \|\boldsymbol{\upsilon}_p\| - \sum_{i=0}^{t-p-1} \left\|A^i(\boldsymbol{\phi})\right\| |D_{t-i}(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))| \\ & + \left|\sum_{i=0}^{t-p-1} A_{1,1}^i(\boldsymbol{\phi})(1-\sum_{i=1}^{p}\phi_i)\mu\right| - |y_t|\,. \end{aligned} \tag{83}$$

From Lemma 4 there exists a positive constant C_1 and $0 < \nu < 1$ such that for $t \geq p+1$

$$\begin{aligned} & \left\|A(\boldsymbol{\phi})^{t-p}\right\| \|\boldsymbol{\upsilon}_p\| + \sum_{i=0}^{t-p-1} \left\|A^i(\boldsymbol{\phi})\right\| |D_{t-i}(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))| \\ \leq \quad & C_1 \left(\nu^t \|\boldsymbol{\upsilon}_p\| + \sum_{i=0}^{t-p-1} \nu^i |D_{t-i}(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))| \right). \end{aligned}$$

Since $\mathcal{B}_0$ is compact, η is bounded, and according to (19), $0 < \widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}) \leq \widehat{\sigma}_y$, we have that there exists a positive constant C_2 such that for $t \geq p+1$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} |D_t(\boldsymbol{\beta},\sigma)| \leq C_2\widehat{\sigma}_y$$

and then there exists a positive constant C such that

$$\begin{aligned} & \sup_{\boldsymbol{\beta}\in\mathcal{B}} \left\|A(\boldsymbol{\phi})^{t-t_0}\right\| \|\boldsymbol{\upsilon}_p\| + \sum_{i=0}^{t-p-1} \left\|A^i(\boldsymbol{\phi})\right\| \left|D_{t-i}(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))\right| \\ \leq\ & C\left(\widehat{\sigma}_y + \nu^t \sum_{t=1}^{p} |y_t|\right). \end{aligned} \tag{84}$$

Define for $t \geq p+1$,

$$\xi_t(\boldsymbol{\phi},\mu) = \sum_{i=1}^{p} \phi_i \xi_{t-i}(\boldsymbol{\phi},\mu) + (1 - \sum_{i=1}^{p} \phi_i)\mu$$

with $\xi_p = 0, \xi_{p-1} = 0, \cdots, \xi_1 = 0$. It is easy to show that

$$\xi_t(\boldsymbol{\phi},\mu) = (1 - \sum_{i=1}^{p} \phi_i)\mu \sum_{i=0}^{t-p-1} A_{1,1}^i(\boldsymbol{\phi}).$$

Then, from Lemma 3 and using an argument similar to the one used to prove (59) in Lemma 6, there exists $\varepsilon > 0$ and t_0 such that for all $t \geq t_0$,

$$\inf_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} |\xi_t(\boldsymbol{\phi},\mu)| \geq \frac{\varepsilon}{2} |\mu|. \tag{85}$$

Then, from (83), (84) and (85), for all $t \geq t_0$,

$$\inf_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \left|a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})\right| \geq \frac{\varepsilon}{2} |\mu| - |y_t| - C\left(\widehat{\sigma}_y + \nu^t \sum_{t=1}^{p} |y_t|\right). \tag{86}$$

Since $\sup \rho_1 > b$ and $\lim_{x\to\infty} \rho_1(|x|) = \sup \rho_1$, there exists k_0 and $\lambda > 1$ such that for all $|x| \geq k_0$,

$$\rho_1(x) \geq \lambda b. \tag{87}$$

Since $\lim_{n\to\infty} \widehat{\sigma}_y = \sigma_y$ a.s. and $\lim_{t\to\infty} \nu^t \sum_{i=1}^{p} |y_i| = 0$ a.s., with probability one there exists $t_1 \geq t_0$ such that for all $t \geq t_1$, $\widehat{\sigma}_y + \nu^t \sum_{i=1}^{p} |y_i| \leq 2\sigma_y$. Take k_1 such that the set

$$R_t = \{|y_t| \leq k_1 - 2C\sigma_y\} \tag{88}$$

satisfies $P(R_t) \geq 1/\lambda$.Take $k = \max(k_1/(s_0+1), k_0)$ and d such that $d > 4k(s_0+1)/\varepsilon$. Then, from the definition of k, (86) and (88) in R_t, for all $t \geq t_1$

$$\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \left|a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))\right| > k(s_0+1). \tag{89}$$

Since $\rho_1$satisfies property P1, we have,

$$\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\frac{a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))}{s_0+1}\right)$$
$$\geq \frac{1}{n-p}\sum_{t=t_1+1}^{n}\rho_1\left(\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\left|\frac{a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))}{s_0+1}\right|\right) I_{R_t}. \quad (90)$$

From (87) and (89) and since R_t is stationary and ergodic with $P(R_t) \geq 1/\lambda$ we get

$$\liminf_{n\to\infty} \frac{1}{n-p}\sum_{t=t_1+1}^{n}\rho_1\left(\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\left|\frac{a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))}{s_0+1}\right|\right) I_{R_t}$$
$$\geq \liminf_{n\to\infty} \frac{\rho_1(k)}{n-p}\sum_{t=t_1+1}^{n} I_{R_t}$$
$$= \liminf_{n\to\infty} \frac{\rho_1(k)}{n-p}\sum_{t=p+1}^{n} I_{R_t} - \lim_{n\to\infty}\frac{\rho_1(k)}{n-p}(t_1-p) \geq b \text{ a.s.}$$

and then from (90) we have,

$$\liminf_{n\to\infty}\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} \frac{1}{n-p}\sum_{t=p+1}^{n}\rho_1\left(\frac{a_t^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta}))}{s_0+1}\right) \geq b \text{ a.s..}$$

This proves the Lemma.

Proof of Theorem 2.

From Lemmas 8 and 9 we have that there exists $\delta > 0$ such that

$$\liminf_{n\to\infty}\inf_{\boldsymbol{\beta}\in\mathcal{B}} S_n(\mathbf{a}_n^b(\boldsymbol{\beta},\widehat{\sigma}(\boldsymbol{\phi},\boldsymbol{\theta})\ ) > s_0+\delta \text{ a.s..}$$

But, by Theorem 1-(ii) we have that $\widehat{\boldsymbol{\beta}}_S$ satisfy

$$\lim_{n\to\infty} S_n(\mathbf{a}_n(\widehat{\boldsymbol{\beta}}_S)) = s_0 \text{ a.s..}$$

This proves the Theorem.

The following four Lemmas will be used to prove Theorem 3.

Lemma 10 *Assume that y_t satisfies condition P2 with innovations satisfying P3 and assume that ρ_2 satisfy condition P1 with ρ_2 bounded. Let us call $m(\boldsymbol{\beta}) = E(\rho_2(a_t^e(\boldsymbol{\beta})/s_0))$, then*

$$\boldsymbol{\beta}_0 = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} m(\boldsymbol{\beta}).$$

Proof. Similar to Lemma 2.

Lemma 11 *Assume that y_t satisfies condition P2 and ρ_2 condition P1. Define*

$$M_n^e(\boldsymbol{\beta}) = \frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right). \tag{91}$$

Then, we have,

$$\lim_{n\to\infty}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left|M_n^e(\boldsymbol{\beta}) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right)\right| = 0 \text{ a.s.}$$

for all $d > 0$.

Proof.

By the dominated convergence theorem, it is easy to show that

$$M(\boldsymbol{\beta}, v) = E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{v}\right)\right)$$

is continuous. Then, given $\epsilon > 0$, there exists δ with $0 < \delta < s_0$ such that

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\nu\in[s_0-\delta,s_0+\delta]}\left|E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{v}\right)\right) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right)\right| < \frac{\epsilon}{2}. \tag{92}$$

Since $a_t^e(\boldsymbol{\beta})$ is stationary and ρ_2 continuous and bounded, by Lemma 3 of Muler and Yohai [22] we have

$$\begin{aligned}&\lim_{n\to\infty}\sup_{(\boldsymbol{\beta},v)\in\mathcal{C}_0}\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right)\right)\right| \\ &= \quad 0 \text{ a.s.,}\end{aligned} \tag{93}$$

where

$$\mathcal{C}_0 = \{(\boldsymbol{\beta},v):\ \beta\in\mathcal{B}_0\times[-d,d],\nu\in[s_0-\delta,s_0+\delta]\}$$

By Theorem 2, $\lim_{n\to\infty} s_n^* = s_0$ a.s.. Then, with probability one there exists n_0 such that for all $n \geq n_0$ we have $s_n^* \in [s_0-\delta, s_0+\delta]$ and

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\nu\in[s_0-\delta,s_0+\delta]}\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right)\right)\right| < \frac{\epsilon}{2}. \tag{94}$$

Hence, from (92) and (94) we have that for $n \geq n_0$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right)\right| \leq$$

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\nu\in[s_0-\delta,s_0+\delta]}\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right) - E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{\nu}\right)\right)\right|$$

$$+\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d],\nu\in[s_0-\delta,s_0+\delta]}\left|E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}}{\nu}\right)\right)-E\left(\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}}{s_0}\right)\right)\right|<\varepsilon.$$

This proves the Lemma.

Lemma 12. *Under the assumptions of Theorem 3, we have*

$$\lim_{n\to\infty}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}|M_n(\boldsymbol{\beta})-M_n^e(\boldsymbol{\beta})|=0\text{ a.s.}.$$

Proof.

We have

$$M_n(\boldsymbol{\beta})-M_n^e(\boldsymbol{\beta})=\frac{1}{n-p}\sum_{t=p+1}^{n}\left(\rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right).$$

From (54), since ρ_2' is bounded and $\lim_{n\to\infty}s_n^*=s_0>0$ a.s. there exists $k>0$, $0<\nu<1$, a random variable Z and n_0 such that for all $n\geq n_0$

$$\left|\rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right|\leq k\left|a_t(\boldsymbol{\beta})-a_t^e(\boldsymbol{\beta})\right|\leq k\nu^tZ. \tag{95}$$

Then,

$$\begin{aligned}&\lim_{n\to\infty}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\frac{1}{n-p}\left|\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right|\\ \leq\ &\lim_{n\to\infty}\frac{kZ}{n-p}\sum_{t=p+1}^{n}\nu^t\leq\lim_{n\to\infty}\frac{kZ}{(n-p)(1-\nu)}=0.\end{aligned} \tag{96}$$

This proves the Lemma.

Lemma 13. *Under the assumptions of Theorem 3, there exists $d>0$ and $\delta>0$ such that*

$$\liminf_{n\to\infty}\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}M_n(\boldsymbol{\beta})\geq m(\boldsymbol{\beta}_0)+\delta\text{ a.s.},$$

where $m(\boldsymbol{\beta}_0)$ is defined in Lemma 10.

Proof.

Since the innovation a_t satisfy P3, then $m(\boldsymbol{\beta}_0)=E\left(\rho_2\left(a_t/s_0\right)\right)<\sup\rho_2$. Since $\lim_{x\to\infty}\rho_2(|x|)=\sup\rho_2$, using similar arguments to those used in Lemma 6, we have that there exists $d>0$ and $\lambda>1$ such that

$$\liminf_{n\to\infty}\inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0}\frac{1}{n-p}\sum_{t=p+1}^{n}\rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)>\lambda m(\boldsymbol{\beta}_0)\text{ a.s.}. \tag{97}$$

and so the Lemma follows.

Proof of Theorem 3. Follows from Lemmas 10-13 using similar arguments as those used in the proof of Theorem 1.

The next two Lemmas will be used to prove Theorem 4.

Lemma 14. *Under the assumptions of Theorem 3, for all $d > 0$ there exists $\delta > 0$ such that*

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} M_n^b(\boldsymbol{\beta}) \geq m(\boldsymbol{\beta}_0) + \delta \quad \text{a.s.},$$

where $m(\boldsymbol{\beta}_0)$ is defined in Lemma 10.

Proof. It is similar to the proof of Lemma 8.

Lemma 15. *Under the assumptions of Theorem 3 , there exists $d > 0$ and $\delta > 0$ such that*

$$\liminf_{n\to\infty} \inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} M_n^b(\boldsymbol{\beta}) \geq m(\boldsymbol{\beta}_0) + \delta \text{ a.s.},$$

where $m(\boldsymbol{\beta}_0)$ is defined as in Lemma 10.

Proof.

Since the innovations a_t satisfy P3, then $m(\boldsymbol{\beta}_0) = E\left(\rho_2\left(a_t/s_0\right)\right) < \sup \rho_2$.Since $\lim_{x\to\infty} \rho_2(|x|) = \sup \rho_2$, using similar arguments to those used in Lemma 9 we have that there exists $d > 0$ and $\lambda > 1$,

$$\liminf_{n\to\infty} \inf_{|\mu|>d,(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathcal{B}_0} M_n^b(\boldsymbol{\beta}) \ > \lambda m(\boldsymbol{\beta}_0) \text{ a.s..} \tag{98}$$

and so the Lemma follows.

Proof of Theorem 4

From Lemmas 14 and 15 we have that there exists $\delta > 0$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\beta}\in\mathcal{B}} M_n^b(\boldsymbol{\beta}) \geq m(\boldsymbol{\beta}_0) + \delta.$$

Theorem 3 implies that $\lim_{n\to\infty} M_n(\widehat{\boldsymbol{\beta}}_M) = m(\boldsymbol{\beta}_0)$ a.s.. This proves the Theorem.

The next five Lemmas will be used to prove Theorem 5.

Lemma 16. *Under the assumptions of Theorem 5, we have*

$$\frac{1}{(n-p)^{1/2}} \sum_{t=p+1}^{n} \nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_\mathbf{0})}{s_0}\right) \to_D N\left(\mathbf{0}, V_0\right),$$

where

$$V_0 = E\left(\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_\mathbf{0})}{s_0}\right)\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_\mathbf{0})}{s_0}\right)'\right). \tag{99}$$

Proof.

We can write

$$\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right) = \frac{\psi_2(a_t/s_0)}{s_0}\nabla a_t^e(\boldsymbol{\beta}_0). \tag{100}$$

Since ψ_2 is odd and the distribution of a_t is symmetric,

$$E\left(\psi_2\left(\frac{a_t}{s_0}\right)\right) = 0. \tag{101}$$

From Lemma 1 (ii) and the fact that $E(y_t^2) < \infty$ we get

$$V_0 = E(\nabla a_t^e(\boldsymbol{\beta}_0)\nabla a_t^e(\boldsymbol{\beta}_0)') < \infty.$$

Therefore, from (100) and (101) and since $\nabla a_t^e(\boldsymbol{\beta})$ depends only on

$$\mathbf{Y_{t-1}} = (y_{t-1}, y_{t-2}, ...)$$

for any column vector $\mathbf{c} \neq \mathbf{0}$ in R^{p+q+1} we have that

$$\mathbf{c}'\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right) = \mathbf{c}'\nabla\rho\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right)$$

is a stationary martingale difference sequence. Then, applying the Central Limit Theorem for Martingales (see Theorem 24.3, Davidson [7]) we have that

$$\frac{1}{(n-p)^{1/2}}\sum_{t=p+1}^{n}\mathbf{c}'\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right) \to_D N(0, \mathbf{c}'V_0\mathbf{c}).$$

This implies that

$$\frac{1}{(n-p)^{1/2}}\sum_{t=p+1}^{n}\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right) \to_D N(0, V_0)$$

proving the Lemma.

Lemma 17. *Under the assumptions of Theorem 5 we have*

$$\lim_{n\to\infty}\frac{1}{(n-p)^{1/2}}\left\|\sum_{t=p+1}^{n}\left(\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_n^*}\right) - \nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right)\right)\right\| = 0$$

in probability.

Proof.

The proof is similar to the one of Lemma 5.1 in Yohai [25] for MM-estimates in the case of regression. We can write,

$$\begin{aligned} & \frac{1}{(n-p)^{1/2}}\sum_{t=p+1}^{n}\left(\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_n^*}\right) - \nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0}\right)\right) \qquad (102) \\ = \; & \frac{1}{(n-p)^{1/2}}\sum_{t=p+1}^{n}\left(\frac{\psi_2(a_t^e/s_n^*)}{s_n^*} - \frac{\psi_2(a_t^e/s_0)}{s_0}\right)\nabla a_t^e(\boldsymbol{\beta}_0). \end{aligned}$$

Define for $0 \leq v \leq 1$, $1 \leq j \leq p+q+1$,

$$A_{n,j}(v) = \frac{1}{(n-p)^{1/2}} \sum_{t=p+1}^{n} \psi_2 \left(\frac{a_t}{(0.5+v)\, s_0} \right) \nabla_j a_t^e(\boldsymbol{\beta}_\mathbf{0}).$$

Since from Theorem 2 $\lim_{n\to\infty} s_n^* = s_0$ a.s. in order to prove the Lemma is enough to show that $A_{n,j}(v)$ $1 \le j \le p+q+1$, are tight. Using Theorem 12.3 of Billingsley [2] it will be enough to show the following two conditions,

(i) $A_{n,j}(0)$ is tight.

(ii)For any $0 \le v_1 \le v_2$ and any $\lambda > 0$ we have that there exists a constant k_1 such that

$$P(|A_{n,j}(v_2) - A_{n,j}(v_1)| \ge \lambda) \le \frac{k_1}{\lambda^2}(v_2 - v_1)^2.$$

(i) follows from Lemma 16.

Let us prove now (ii). We can write for $1 \le j \le p+q+1$,

Put

$$G(a,v) = \psi_2 \left(\frac{a}{(0.5+v)\, s_0} \right).$$

Then

$$\begin{aligned} & E\left((A_{n,j}(v_2) - A_{n,j}(v_1))^2\right) \\ = \ & \frac{1}{n-p} E\left(\left(\sum_{t=p+1}^{n} (G(a_t, v_2) - G(a_t, v_1)) \nabla_j a_t^e(\boldsymbol{\beta}_\mathbf{0}) \right)^2 \right) \\ = \ & \frac{1}{n-p} \sum_{t=p+1}^{n} \sum_{r=p+1}^{n} E(B_r C_r B_t C_t), \end{aligned} \tag{103}$$

where

$$B_t = \psi_2 \left(\frac{a_t}{(0.5+v_2)\, s_0} \right) - \psi_2 \left(\frac{a_t}{(0.5+v_1)\, s_0} \right) \tag{104}$$

and

$$C_t = \nabla_j a_t^e(\boldsymbol{\beta}_\mathbf{0}). \tag{105}$$

Since $\nabla a_t^e(\boldsymbol{\beta}_0)$ depends on $\mathbf{Y_{t-1}}$ =$(y_{t-1}, y_{t-2}, ...)$, we have that B_t is independent of C_r for all $r \le t$. Moreover all B_t 's are independent. Then if $r < t$, using that $E(B_t|Y_{t-1}) = E(B_t) = 0$, we obtain

$$\begin{aligned} E(B_r B_t C_r C_t) &= E(E(B_r B_t C_r C_t | \mathbf{Y}_{t-1})) \\ &= E(E(B_t|Y_{t-1}) C_t C_r B_r) \\ &= 0. \end{aligned} \tag{106}$$

Moreover

$$E(B_t^2 C_t^2) = E(B_t^2) E(C_t^2). \tag{107}$$

From (103), (104), (105), (106) and (107) we obtain

$$
\begin{aligned}
& E\left((A_{n,j}(v_2)-A_{n,j}(v_1))^2\right) \\
= \; & E\left(\psi_2\left(\frac{a_t}{(0.5+v_2)\,s_0}\right)-\psi_2\left(\frac{a_t}{(0.5+v_1)\,s_0}\right)\right)^2 E(\nabla_j a_t^e(\boldsymbol{\beta_0}))^2. \quad (108)
\end{aligned}
$$

Let $v_1 < v < v_2$.Then, using the Mean Value Theorem we get

$$
\begin{aligned}
& E\left(\psi_2\left(\frac{a_t}{(0.5+v_2)\,s_0}\right)-\psi_2\left(\frac{a_t}{(0.5+v_1)\,s_0}\right)\right)^2 \\
= \; & \frac{(v_2-v_1)^2}{s_0^2\,(0.5+v)^4} E\left(\left(a_t\psi_2'\left(\frac{a_t}{(0.5+v)\,s_0}\right)\right)^2\right).
\end{aligned}
$$

Then, since ψ_2' is bounded, a_t has second moment and $s_0 > 0$ we can conclude that there exists $k_0 > 0$ such that

$$
E\left(\psi_2\left(\frac{a_t}{(0.5+v_2)\,s_0}\right)-\psi_2\left(\frac{a_t}{(0.5+v_1)\,s_0}\right)\right)^2 \leq k_0\,(v_2-v_1)^2. \qquad (109)
$$

Then since $E(y_t^2) < \infty$, by Lemma 1 (ii) we have that $E\left((\nabla_j a_t^e(\boldsymbol{\beta}_0))^2\right) < \infty$. Then from (108) and (109) there exists $k_1 > 0$ such that

$$
E\left((A_{n,j}(v_2)-A_{n,j}(v_1))^2\right) \leq k_1\,(v_2-v_1)^2.
$$

Hence, (ii) follows from the Chebyshev's inequality.

Lemma 18. *Under the assumptions of Theorem 5, we have*

$$
\lim_{n\to\infty} \frac{1}{(n-p)^{1/2}} \left\| \sum_{t=p+1}^{n} \left(\nabla\rho_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right)-\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta_0})}{s_n^*}\right)\right)\right\| = 0 \text{ a.s..}
$$

Proof.
We can write

$$
\begin{aligned}
& \nabla\rho_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right)-\nabla\rho_2\left(\frac{a_t^e(\boldsymbol{\beta_0})}{s_n^*}\right) \\
= \; & \frac{1}{s_n^*}\left(\psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right)\nabla a_t(\boldsymbol{\beta_0})-\psi_2\left(\frac{a_t}{s_n^*}\right)\nabla a_t^e(\boldsymbol{\beta_0})\right) \\
= \; & \frac{1}{s_n^*}\psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right)(\nabla a_t(\boldsymbol{\beta_0})-\nabla a_t^e(\boldsymbol{\beta_0})) \\
& +\frac{1}{s_n^*}\left(\psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right)-\psi_2\left(\frac{a_t}{s_n^*}\right)\right)\nabla a_t^e(\boldsymbol{\beta_0}). \qquad (110)
\end{aligned}
$$

By Lemma 1 (ii), (29), (30) and (31), using similar arguments to those leading to the proof of (54) in Lemma 5, we can prove that there exists $0 < \upsilon < 1$ and a random variable W such that

$$\|\nabla a_t(\boldsymbol{\beta}_0) - \nabla a_t^e(\boldsymbol{\beta}_0)\| \leq \upsilon^t W \tag{111}$$

and therefore

$$\lim_{n\to\infty} \frac{1}{(n-p)^{1/2}} \sum_{t=p+1}^{n} \|\nabla a_t(\boldsymbol{\beta_0}) - \nabla a_t^e(\boldsymbol{\beta_0})\| = 0 \text{ a.s..}$$

Then, since ψ_2 is bounded and by Theorem 2 (ii) we have $\lim_{n\to\infty} s_n^* = s_0 > 0$ a.s.,

$$\lim_{n\to\infty} \frac{1}{(n-p)^{1/2} s_n^*} \left\| \sum_{t=p+1}^{n} \psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right) (\nabla a_t(\boldsymbol{\beta_0}) - \nabla a_t^e(\boldsymbol{\beta_0})) \right\| = 0 \text{ a.s..} \tag{112}$$

Using that ψ_2' is bounded and the Mean Value Theorem we can found a constant $k_1 > 0$ such that

$$\begin{aligned} &\frac{1}{(n-p)^{1/2} s_n^*} \sum_{t=p+1}^{n} \left\| \left(\psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right) - \psi_2\left(\frac{a_t}{s_n^*}\right)\right) \nabla a_t^e(\boldsymbol{\beta_0}) \right\| \\ \leq \quad & \frac{k_1}{(n-p)^{1/2} s_n^*} \sum_{t=p+1}^{n} |a_t(\boldsymbol{\beta_0}) - a_t^e(\boldsymbol{\beta_0})| \, \|\nabla a_t^e(\boldsymbol{\beta_0})\| \end{aligned} \tag{113}$$

From (54), there exists $0 < \nu < 1$ and a random variable Z such that

$$\sum_{t=p+1}^{n} |a_t(\boldsymbol{\beta_0}) - a_t^e(\boldsymbol{\beta_0})| \, \|\nabla a_t^e(\boldsymbol{\beta_0})\| \leq Z \sum_{t=p+1}^{n} \nu^t \left(\|\nabla a_t^e(\boldsymbol{\beta_0})\|\right). \tag{114}$$

From (29)-(31) we have that $W_1 = \sum_{t=p+1}^{n} \nu^t \left(\|\nabla a_t^e(\boldsymbol{\beta_0})\|\right)$ is well defined. Then, from (113), (114) and the fact that $\lim_{n\to\infty} s_n^* = s_0$ a.s., we obtain

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{(n-p)^{1/2} s_n^*} \sum_{t=p+1}^{n} \left\| \left(\psi_2\left(\frac{a_t(\boldsymbol{\beta_0})}{s_n^*}\right) - \psi_2\left(\frac{a_t}{s_n^*}\right)\right) \nabla a_t^e(\boldsymbol{\beta_0}) \right\| \\ \leq \quad & \lim_{n\to\infty} \frac{k_1 Z W_1}{(n-p)^{1/2} s_n^*} = 0 \text{ a.s..} \end{aligned} \tag{115}$$

Then the Lemma follows from (110), (112) and (115).

Lemma 19. *Under the assumptions of Theorem 5 we have for all $d > 0$,*
(i)

$$\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \left\| \frac{1}{n-p} \sum_{t=p+1}^{n} \nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right) - E\left(\nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_0}\right)\right) \right\| = 0 \text{ a.s.,}$$

where $||A||$ denotes the l_2 norm of the matrix A.
(ii)

$$E\left(\nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta_0})}{s_0}\right)\right) = \frac{1}{s_0^2}E\left(\psi_2'\left(\frac{a_t}{s_0}\right)E\left(\nabla a_t^e(\boldsymbol{\beta_0})\nabla a_t^e(\boldsymbol{\beta_0})'\right)\right)$$

and this matrix is non-singular.
Proof.
The proof of (i) is similar to the one of Lemma 11.
We now prove (ii). Using that $E(\psi_2(a_t/s_0)) = 0$ and that, according to (29)-(37), both $\nabla a_t^e(\boldsymbol{\beta})$ and $\nabla^2 a_t^e(\boldsymbol{\beta})$ depend on $\mathbf{Y_{t-1}} = (y_{t-1}, y_{t-2}, ...)$, we obtain

$$E\left(\nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta_0})}{s_0}\right)\right) = \frac{1}{s_0^2}E\left(\psi_2'\left(\frac{a_t}{s_0}\right)\right)E\left(\nabla a_t^e(\boldsymbol{\beta_0})\nabla a_t^e(\boldsymbol{\beta_0})'\right).$$

Since $E(\psi_2'(a_t/s_0)) > 0$ and $E\left(\nabla a_t^e(\boldsymbol{\beta_0})\nabla a_t^e(\boldsymbol{\beta_0})'\right)$ is a non-singular matrix (see Bustos and Yohai [4]) we obtain (ii).

Lemma 20. *Under the assumptions of Theorem 5, we have,*

$$\lim_{n\to\infty}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\frac{1}{n-p}\left\|\sum_{t=p+1}^{n}\left(\nabla^2\rho_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right)\right\| = 0 \text{ a.s.}$$

for all $d > 0$.
Proof.
Put

$$V(\boldsymbol{\beta}) = \nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'.$$

Differentiating $\nabla\rho\left(a_t^e(\boldsymbol{\beta})/s_n^*\right)$ we obtain

$$\nabla^2\rho_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right) = \frac{\psi_2(a_t^e(\boldsymbol{\beta})/s_n^*)}{s_n^*}\nabla^2 a_t^e(\boldsymbol{\beta}) + \frac{\psi_2'(a_t^e(\boldsymbol{\beta})/s_n^*)}{(s_n^*)^2}V(\boldsymbol{\beta}). \tag{116}$$

Let us define

$$G_t(\boldsymbol{\beta}) = \frac{1}{(s_n^*)^2}\left(\psi_2'\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})' - \psi_2'\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right)$$

and

$$H_t(\boldsymbol{\beta}) = \frac{1}{s_n^*}\left(\psi_2\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)\nabla^2 a_t(\boldsymbol{\beta}) - \psi_2\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\nabla^2 a_t^e(\boldsymbol{\beta})\right),$$

and then

$$\nabla^2\rho\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right) - \nabla^2\rho\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right) = G_t(\boldsymbol{\beta}) + H_t(\boldsymbol{\beta}). \tag{117}$$

We can write

$$\begin{aligned} G_t(\boldsymbol{\beta}) &= \frac{1}{(s_n^*)^2}\psi_2'\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)\left(\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})'-\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right) \\ &\quad +\frac{1}{(s_n^*)^2}\left(\psi_2'\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\psi_2'\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right)\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'. \end{aligned}$$

We also have

$$\begin{aligned} & \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left\|\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})'-\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right\| \\ \leq & \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left(\|\nabla a_t(\boldsymbol{\beta})+\nabla a_t^e(\boldsymbol{\beta})\|\right)\left(\|\nabla a_t(\boldsymbol{\beta})-\nabla a_t^e(\boldsymbol{\beta})\|\right). \end{aligned}$$

Then from Lemma 1 (ii) ,(29)-(31) , using similar arguments to those leading (54) in Lemma 5 , we can prove that there exist $0<\upsilon<1$ and a random variable W such that for $t\geq p+1$

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\|\nabla a_t(\boldsymbol{\beta})-\nabla a_t^e(\boldsymbol{\beta})\|\leq\upsilon^t W,\ \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\|\nabla a_t^e(\boldsymbol{\beta})\|\leq W.$$

Then, putting $V=3W^2$ we have for $t\geq p+1$,

$$\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\left\|\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})'-\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right\|\leq\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\nu^t V$$

and then

$$\lim_{n\to\infty}\frac{1}{n-p}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\sum_{t=p+1}^{n}\left\|\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})'-\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right\|=0$$

a.s..

Hence, since ψ_2' is bounded and $\lim_{n\to\infty}s_n^*=s_0>0$ a.s. we have

$$\begin{aligned} & \lim_{n\to\infty}\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]}\sum_{t=p+1}^{n}\left\|\frac{\psi_2'\left(a_t(\boldsymbol{\beta})/s_n^*\right)}{(s_n^*)^2(n-p)}\left(\nabla a_t(\boldsymbol{\beta})\nabla a_t(\boldsymbol{\beta})'-\nabla a_t^e(\boldsymbol{\beta})\nabla a_t^e(\boldsymbol{\beta})'\right)\right\| \\ = & \ 0 \text{ a.s..} \end{aligned} \tag{118}$$

Since ψ_2'' is bounded , there exists a constant $k_1>0$ such that

$$\left|\psi_2'\left(\frac{a_t(\boldsymbol{\beta})}{s_n^*}\right)-\psi_2'\left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*}\right)\right|\leq\frac{k_1}{s_n^*}\left|a_t(\boldsymbol{\beta})-a_t^e(\boldsymbol{\beta})\right|.$$

Therefore, from (54), we have that there exists $0<\nu<1$ and a random variable Z such that for all $t\geq p+1$,

$$
\begin{aligned}
&\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sum_{t=p+1}^{n} \left\| \left(\psi_2' \left(\frac{a_t(\boldsymbol{\beta})}{s_n^*} \right) - \psi_2' \left(\frac{a_t^e(\boldsymbol{\beta})}{s_n^*} \right) \right) V(\boldsymbol{\beta}) \right\| \\
\leq \quad & \frac{k_1 Z}{s_n^*} \sum_{t=p+1}^{n} \nu^t \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \|V(\boldsymbol{\beta})\ \| \qquad (119)
\end{aligned}
$$

and since by Lemma 1-(ii), $E(y_t^2) < \infty$ we have

$$
\sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} E\left(\|V(\boldsymbol{\beta})\|\right) < \infty
$$

and then

$$
\lim_{n\to\infty} \frac{1}{n-p} \sum_{t=p+1}^{n} \nu^t \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \|V(\boldsymbol{\beta})\ \| = 0 \text{ a.s..} \qquad (120)
$$

Using that $\lim_{n\to\infty} s_n^* = s_0$ a.s., from (119) and (120) we obtain

$$
\begin{aligned}
&\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \sum_{t=p+1}^{n} \left\| \left(\frac{\psi_2'\left(a_t(\boldsymbol{\beta})/s_n^*\right) - \psi_2'\left(a_t^e(\boldsymbol{\beta})/s_n^*\right)}{\left(s_n^*\right)^2 (n-p)} \right) V(\boldsymbol{\beta}) \right\| \\
= \quad & 0 \text{ a.s..} \qquad (121)
\end{aligned}
$$

Therefore, from (118), (121) we have that

$$
\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \frac{1}{n-p} \sum_{t=p+1}^{n} \|G_t(\boldsymbol{\beta})\| = 0 \text{ a.s..} \qquad (122)
$$

Similarly, using (32)-(37) and Lemma 1 (iii) we can prove

$$
\lim_{n\to\infty} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \frac{1}{n-p} \sum_{t=p+1}^{n} \|H_t(\boldsymbol{\beta})\| = 0 \text{ a.s..} \qquad (123)
$$

Then the Lemma follows from (117), (122) and (123).

Proof of Theorem 5.

The estimate $\widehat{\boldsymbol{\beta}}_M$ satisfies

$$
\sum_{t=p+1}^{n} \nabla\rho_2 \left(\frac{a_t(\widehat{\boldsymbol{\beta}}_M)}{s_n^*} \right) = 0.
$$

Then, using the Mean Value Theorem we have

$$
\sum_{t=p+1}^{n} \nabla\rho_2 \left(\frac{a_t(\boldsymbol{\beta}_0)}{s_n^*} \right) + \left(\sum_{t=p+1}^{n} \nabla^2\rho_2 \left(\frac{a_t(\boldsymbol{\beta}^*)}{s_n^*} \right) \right) (\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0) = 0, \qquad (124)
$$

where $\boldsymbol{\beta}^*$ is an intermediate point between $\widehat{\boldsymbol{\beta}}_M$ and $\boldsymbol{\beta}_0$.

From Theorem 3 we have that $\widehat{\boldsymbol{\beta}}_M \rightarrow \boldsymbol{\beta}_0$ a.s.. Take $d > 0$ so that $d > |\mu_0|$, then, with probability one there exists n_0 such that $\widehat{\boldsymbol{\beta}}_M \in \mathcal{B}_0 \times [-d, d]$ for all $n \geq n_0$. From Lemmas 19 (i) and 20 we get

$$\begin{aligned} & \lim_{n\rightarrow\infty} \sup_{\boldsymbol{\beta}\in\mathcal{B}_0\times[-d,d]} \left\| \frac{1}{n-p} \sum_{t=p+1}^{n} \left(\nabla^2 \rho_2 \left(\frac{a_t(\boldsymbol{\beta})}{s_n^*} \right) - E\left(\nabla^2 \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta})}{s_0} \right) \right) \right) \right\| \\ = \; & 0 \text{ a.s..} \end{aligned} \tag{125}$$

Put

$$A_n = \frac{1}{n-p} \sum_{t=p+1}^{n} \nabla^2 \rho_2 \left(\frac{a_t(\boldsymbol{\beta}^*)}{s_n^*} \right). \tag{126}$$

Then, since $\boldsymbol{\beta}^* \rightarrow \boldsymbol{\beta}_0$ a.s. and $E\left(\nabla^2 \rho_2 \left(a_t^e(\boldsymbol{\beta})/s_0\right)\right)$ is continuous in $\boldsymbol{\beta}$ we have that

$$\lim_{n\rightarrow\infty} A_n = E\left(\nabla^2 \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right) \right) \text{ a.s..} \tag{127}$$

Therefore from Lemma 19 (ii), for n large enough A_n is non singular. Then, from (124) we get for n large enough

$$(n-p)^{1/2}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0) = A_n^{-1}\mathbf{c}_n, \tag{128}$$

where

$$\mathbf{c}_n = \frac{1}{(n-p)^{1/2}} \sum_{t=p+1}^{n} \nabla \rho_2 \left(\frac{a_t(\boldsymbol{\beta}_0)}{s_n^*} \right).$$

From Lemmas 16, 17 and 18 we have

$$\mathbf{c}_n \rightarrow_D N\left(\mathbf{0},\ V_0\right). \tag{129}$$

Then from (127), (128) and (129) we get.

$$(n-p)^{1/2}(\widehat{\boldsymbol{\beta}}_M - \boldsymbol{\beta}_0) \rightarrow_D N\left(\mathbf{0},\ V_1^{-1} V_0 V_1^{-1}\right),$$

where

$$V_1 = E\left(\nabla^2 \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right) \right).$$

From (29), (30) and (31) we have

$$\begin{aligned} \frac{\partial a_t^e(\boldsymbol{\beta}_0)}{\partial \phi_i} &= -\theta_0^{-1}(B)(y_{t-i} - \mu_0) = -\phi_0^{-1} a_{t-i},\ 1 \leq i \leq p, \\ \frac{\partial a_t^e(\boldsymbol{\beta}_0)}{\partial \theta_i} &= \theta_0^{-2}(B)\phi_0(B)(y_{t-i} - \mu) = \theta_0^{-1} a_{t-i},\ 1 \leq i \leq q, \\ \frac{\partial a_t^e(\boldsymbol{\beta}_0)}{\partial \mu} &= -\frac{1 - \sum_{j=1}^{p} \phi_{0j}}{1 - \sum_{j=1}^{p} \theta_{0j}}, \end{aligned}$$

and so we can write

$$\nabla \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right) = \frac{\psi_2 \left(a_t/s_0 \right)}{s_0} \mathbf{v}_t, \tag{130}$$

where $\mathbf{v}_t$ is the stationary process vector of dimension $(p+q+1)$ defined by

$$\mathbf{v}_{tj} = \begin{cases} -\phi_0^{-1} a_{t-j} & \text{if} \quad 1 \le j \le p \\ \theta_0^{-1} a_{t-j-p} & \text{if} \quad p+1 \le j \le p+q \\ \zeta_0 & \text{if} \quad j = p+q+1 \end{cases}$$

where $\zeta_0 = -1 - \sum_{j=1}^{p} \phi_{0j}/(1 - \sum\limits_{j=1}^{p} \theta_{0j})$. Then

$$E \left(\nabla \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right) \nabla \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right)' \right) = \frac{E(\psi_2^2 \left(a_t/s_0 \right))}{s_0^2} E(\mathbf{v}_t \mathbf{v}_t'). \tag{131}$$

Differentiating $\nabla \rho \left(a_t^e(\boldsymbol{\beta})/s_0 \right)$ we obtain

$$\nabla^2 \rho_2 \left(\frac{a_t^e(\boldsymbol{\beta}_0)}{s_0} \right) = \frac{1}{s_0^2} \psi_2' \left(\frac{a_t}{s_0} \right) \mathbf{v}_t \mathbf{v}_t' + \frac{1}{s_0} \psi_2 \left(\frac{a_t}{s_0} \right) \nabla^2 a_t^e(\boldsymbol{\beta}_0). \tag{132}$$

Since $\nabla^2 a_t^e(\boldsymbol{\beta}_0)$ is independent of a_t we have

$$E \left(\psi_2 \left(\frac{a_t}{s_0} \right) \nabla^2 a_t^e(\boldsymbol{\beta}_0) \right) = E \left(\psi_2 \left(\frac{a_t}{s_0} \right) \right) E \left(\nabla^2 a_t^e(\boldsymbol{\beta}_0) \right) = 0$$

and then from (132), since a_t and v_t are independent we get

$$E \left(\nabla^2 \rho_2 \left(\frac{a_t(\boldsymbol{\beta}_0)}{s_0} \right) \right) = \frac{1}{s_0^2} E \left(\psi_2' \left(\frac{a_t}{s_0} \right) \right) E(\mathbf{v}_t \mathbf{v}_t'). \tag{133}$$

Hence, from (131) and (133) we obtain

$$V_1^{-1} V_0 V_1^{-1} = s_0^2 \frac{E(\psi_2^2 \left(a_t/s_0 \right))}{E(\psi_2' \left(a_t/s_0 \right))^2} E \left(\mathbf{v}_t \mathbf{v}_t' \right)^{-1}.$$

Finally is straightforward, see for example Bustos and Yohai [4], to show that

$$E(\mathbf{v}_t \mathbf{v}_t') = \begin{pmatrix} \sigma_a^2 \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \zeta_0^2 \end{pmatrix},$$

where C is defined in the statement of Theorem 5 and $\sigma_a^2 = E(a_t^2)$.

11 REFERENCES

References

[1] Bianco, A. M., Garcia Ben, M. Martinez, E. and Yohai, V. J.(1996). Robust procedures for regression models with ARIMA errors, COMPSTAT 96, *Proceeding of Computational Statistics.* Ed. Albert Prat, pags. 27-38, Physica-Verlag, Heidelberg.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* NewYork. John Wiley and Sons.

[3] Brubacher. S.R. (1974). *Time series outlier detection and modeling with interpolation*, Bell Laboratories Technical Memo.

[4] Bustos, O.H. and Yohai, V.J. (1986), Robust estimates for ARMA models. *J. Amer. Statist. Assoc.* **81**, 155-168.

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

[6] Chen, C. and Liu, L. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series, *J. Amer. Statist. Assoc.* **88**, 284–297.

[7] Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* New York: Oxford University Press.

[8] Davis, R.A., Knight, K. and Liu, J. (1992). M-estimation for autoregression with infinite variance, *Stochastic Process. Appl.*, **40**, 145-180.

[9] Davis, R. A. (1996), Gauss-Newton and M-estimation for ARMA processes with infinite variance, *Stochastic Process. Appl.*, **63**, 75-95.

[10] Fox, A .J. (1972). Outliers in Time Series, J. R. Stat. Soc. Ser. B **34**, 350–363.

[11] Genton, M. G., Lucas, A. (2003). Comprehensive definitions of breakdown-points for independent and dependent observations, J. R. Stat. Soc. Ser. B., **65**, 81-94.

[12] Hampel, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.,* **69**, 383- 393.

[13] Huber, P. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.* **53**, 73-101.

[14] Huber, P. (1981). *Robust Statistics.* Wiley, New York.

[15] Künsch, H. (1984). Infinitesimal robustness for autoregressive processes. *Ann. Statist.,* **12,** 843- 863.

[16] Mancini, L., Ronchetti, E. and Trojani, F. (2005). Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models, *J. Amer. Statist. Assoc.,* **100,** 628-641.

[17] Maronna, R.A. , Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*, Wiley, West Sussex.

[18] Martin, R.D. and Yohai, V.J. (1986). Influence functionals for time series. *Annals of Statistics*, **14,** 781-818.

[19] Martin, R. D. and Yohai, V. (1996). Highly robust estimation of autoregressive integrated time series models. Publicaciones Previas No.89, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

[20] Martin, R.D., Samarov, A. and Vandaele W. (1983). Robust methods for ARIMA models, in Zellner, E. (ed.), *Applied Time Series Analysis of Economic Data*, 153–177, Washington Bureau of the Census.

[21] Masreliez, C.J. (1975). Approximate non-gaussian filtering with linear state and observation relations, *IEEE Trans. Automat. Control* **20**,107-110.

[22] Muler, N. and Yohai V. J. (2002). Robust Estimates for ARCH Processes, *J. Time Ser. Anal.* **23**, 341-375.

[23] Peña, D. (1990). Influential Observations in Time Series, *J. Bus. Econom. Statist.* **8**, 235-341.

[24] Tsay, R.S. (1986). Time series models specification in the presence of outliers, *J. Amer. Statist. Assoc.* **81**, 132–141.

[25] Yohai, V.J. (1985). High breakdown-point and high efficiency robust estimates for regression. Technical Report # 66. Department of Statistics, University of Washington, Seattle, WA.

[26] Yohai, V.J. (1987). High breakdown point and high efficiency estimates for regression, *Ann. Statist.* **15**, 642-65.

[27] Yohai, V.J. and Maronna, R.A. (1977). Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variance, *Ann. Statist.* **5**, 554–560.