Robust Estimates for GARCH Models

By Nora Muler and Victor J. Yohai
Universidad Torcuato Di Tella
and
Universidad de Buenos Aires and CONICET

Abstract

In this paper we present two robust estimates for GARCH(p,q) models. The first is defined by the minimization of a conveniently modified likelihood and the second is similarly defined, but includes an additional mechanism for restricting the propagation of the effect of one outlier on the next estimated conditional variances. We study the asymptotic properties of our estimates proving consistency and asymptotic normality . A Monte Carlo study shows that the quasi maximum likelihood estimate practically collapses when there is a small percentage of outlier contamination, while the proposed robust estimates perform much better. This Monte Carlo study also includes two other robust estimates : a maximum likelihood estimate based on a Student distribution and the least absolute deviation estimate proposed by Peng and Yao. Moreover, we consider several real examples with financial data that illustrate the behavior of these estimates.

Classification code: C22

Keywords: **G**ARCH models, robust estimation, M-estimates, outliers.

1 Introduction

In a seminar paper, Engle (1982) introduced the autoregressive conditional heteroskedastic (ARCH) models. ARCH models were the first of a large family of heteroskedastic time series models such as, for example, the GARCH introduced in Bollerslev (1986), ARCH-M models by Engle, Lilien and Robins (1987) and EGARCH in Nelson (1991).

These models are usually estimated by maximum likelihood assuming that the distribution of one observation conditionally to the past is normal. If the data satisfy the assumption of conditional normality, this procedure is asymptotically efficient. Moreover, even when the conditional distribution of the observations is not normal, these procedures give consistent and asymptotically normal estimates under certain moment conditions. The asymptotic properties of this estimate, known as quasi-maximum likelihood (QML) estimate, were studied by Lee and Hansen (1994) and Lumsdaine (1996) for the GARCH$(1,1)$ model. Elie and Jeantheau (1995) established strong consistency of the QML-estimate in the GARCH $(1,1)$ model and Boussama (2000) proved the asymptotic normality of the same estimate. For the general GARCH(p,q) model, asymptotic properties of this estimator were studied for instance by Berkes, Hováth and Kokoszka (2003), Straumann and Mikosch (2003) and Christian and Zakoïan (2004). These results show that if the innovation has four moments, then the QML–estimate is consistent and has asymptotically normal distribution. Hall and Yao (2003) show that if the fourth moment of the innovation is infinity, then the asymptotic distribution of the QML-estimate may not be normal.

These estimates based on a normal likelihood are very sensitive to the presence of a few outliers in the sample. In fact, a single huge outlier may have a very large effect on the QML-estimate. Estimates that are not much influenced by a small fraction of outliers are called robust estimates.

Several types of outliers have been studied for time series such as additive outliers and innovation outliers. Outliers may be isolated or occur in patches. In our simulated studies we consider only isolated additive outliers. These outliers can be modeled as follows. Suppose that the GARCH(p,q) series is given by x_t. Then, the observed series corresponding to the isolated additive outliers is

$$x_t + v_t u_t,$$

where x_t, v_t and u_t are independent processes. Here v_t, and u_t are sequences of

independent and identically distributed (i.i.d.) random variables and the variable u_t takes values 0 and 1. The event $u_t = 1$ indicates that an outlier occurs at time t, and therefore $x_t + v_t$ is observed instead of x_t. Usually $\theta = P(u_t = 1)$ is small, so that most of the time the GARCH series x_t is observed. Mendes (2000) studied the asymptotic bias produced by additive outliers on the QML-estimate.

Several authors have proposed robust estimates for ARCH models. Sakata and White (1998) proposed estimates based on an M-scale, Mendes and Duarte (1999) defined a class of constrained M-estimates and Muler and Yohai (2002) introduced a class of estimates based on a $\tau-$scale estimate combined with robust filtering. Jiang, Zhao and Hui (2001) proposed L_1 estimates of modified ARCH models. Franses and van Dijk (2000) and Carnero, Peña and Ruiz (2001) used diagnostic procedure for detecting outliers in GARCH models. Rieder, Ruckdeschel and Kohl (2002) introduced a class of robust estimates for a general class of models that includes GARCH, based on the minimization of the mean square error on infinitesimal neighborhoods of contamination. Robust tests for ARCH heteroskedasticity were proposed by van Dijk, Lucas and Franses (1999) and Ronchetti and Trojani (2001).

Li and Kao (2002) proposed a bounded influence estimate for a log GARCH (1,1) model introduced by Geweke (1986). Park (2002) considered a modified GARCH model where the conditional standard deviation (instead of the variances as in GARCH) is modelled as a linear combination of the preceding standard deviations of the absolute values of the preceding observations. The proposed estimator is based on a least absolute deviation (LAD) criterion. Peng and Yao (2003) propose estimates for the GARCH model which are also variations of the LAD criterion. Finally a widespread procedure of protection against heavy tailed distributions in GARCH models uses a maximum likelihood estimate assuming that the conditional distribution given the past is a heavy tailed distribution (like a Stude nt with a small degree of freedom) instead of the normal distribution.

Huber (1981) considers a stricter concept for a robust estimate. It should satisfy the following two properties:

(H1) The estimate should be highly efficient when all observations of the sample follow the assumed model. This condition can be checked by comparing its efficiency to that of the maximum likelihood estimate for that model.

(H2) Replacing a small fraction of observations of the sample by outliers should produce a small change in the estimate. This property was formalized in terms of continuity of the estimate and called *qualitative robustness* by Hampel (1971) for independent observations. Boente, Fraiman and Yohai (1987) generalized this concept for time series.

None of the estimates mentioned above for the GARCH model simultaneously satisfies H1 and H2. If the assumed model is a GARCH model with normal conditional distribution, neither the maximum likelihood estimates corresponding to a heavy tailed distribution nor those based on a LAD criterion satisfy the property H1 stated above. This is shown in Table 1 of Section 4 where we compute the efficiencies of some of these estimate for the normal GARCH. Although these estimators behave much better than the QML-estimate when the conditional distribution has heavy tailed distributions, they also fail to satisfy property H2. See Section 4 where we report of the results of a Monte Carlo study, showing that a small fraction of additive outliers may have a large influence on them.

In this paper we present two classes of robust estimates for GARCH models. The first class can be considered an extension of the M-estimates introduced by Huber (1964) for location and Huber (1973) for regression. They are obtained by maximizing a conveniently modified likelihood function. We show that the M-estimates are consistent and asymptotically normal. These M-estimates are less sensitive to outliers than the QML-estimate and satisfy H1 However they do not satisfy criterion H2, i.e., a few large outliers can still have a large influence on them. This lack of robustness is due to the fact that a single large outlier may have much influence on the conditional variance of an undetermined large number of subsequent observations.

To improve robustness, we propose another estimate that includes an additional mechanism that restricts the propagation of the outlier effect in such a way that the influence of past variances on the present observation are bounded. These estimates are called bounded M-estimates (BM-estimates). BM-estimates are also consistent and asymptotically normal and they possess both properties H1 and H2, i.e., they have a high efficiency under a GARCH normal model and are not much influenced by a small fraction of outlying observations.

In financial data it is very common to use GARCH models to predict stock volatilities, which are one of the parameters required to determine option prices It may be argued that since outlying observations really happen and represent a risk factor, they should be taken into account to determine the option prices. This line of thinking would conclude that for these applications the QML-estimates for GARCH models, which does not downweight the effect of outliers, may be preferable to robust estimates. The following three comments respond to this criticism:

(i) A model with robustly estimated parameters fits the majority of the observations. Instead, if the data contains gross departures from the model, the QML-estimate may fit the bulk of the data poorly while fitting some of the outliers.

(ii) When a robust estimate is used, outliers can be detected as observations not well fitted by the estimated model. This allows the possibility to improve the model by including variables that explain the outliers. However, if a non robust estimate is used, the outliers may remain hidden, thus precluding the possibility of improving the model.

(iii) The downweighting of outliers in the robust estimation process does not preclude their use in prediction, since the estimated coefficients determine only the predictive dynamics of the model. If the recent past contains outliers and the user assumes that these outliers are valid inputs for prediction, they may be used as such without downweighting.

This paper is organized as follows. In Section 2, we state some of the properties of GARCH processes and define the proposed robust estimates. In Section 3, we give the consistency and asymptotically normality results. In Section 4, we report the results of a Monte Carlo Study for the QML -estimate, the Peng and Yao LAD estimator, the maximum likelihood estimates corresponding to a Student distribution with a small degrees of freedom (SML) and our proposed M- and BM-estimators. These results show a clear advantage of the robust estimates when the sample contains outliers, especially in the case of the BM-estimate. In Section 5, we consider examples of series corresponding to daily data and compare the truncated variance and rank correlation of the errors for the daily returns series of the QML-estimate, the LAD-estimate, the SML- and the BM-estimate. Section 6 contains some concluding remarks. Section 7 is an Appendix with some of the proofs. For brevity sake we omit several proofs which can be found in a technical report by Muler and Yohai (2005).

2 Robust Estimates for GARCH(p, q) Models

A series $x_1, ..., x_T$ is a centered GARCH (p, q) process if

$$x_t = \sigma_t z_t, \tag{1}$$

where $z_1, z_2, ..., z_T$ are i.i.d. random variables with a continuous density f such that $E(z_t) = 0$ and $\text{var}(z_t) = 1$($\text{var}(x)$ denotes variance of x) and where the conditional variances σ_t^2 are given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2,$$

where $\alpha_i \geq 0$, $1 \leq i \leq p$, $\beta_i \geq 0, 1 \leq i \leq q$ and $\alpha_0 > 0$. We denote $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_p)$, $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)$ and $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$. When $q = 0$ we obtain the class of ARCH models introduced by Engle (1982).

A necessary and sufficient condition for strict stationarity of the process x_t with finite variance is

$$\Sigma_{i=1}^{p} \alpha_i + \Sigma_{i=1}^{q} \beta_i < 1, \tag{2}$$

see Bollerslev (1986), Nelson (1990), Bougerol and Picard (1992) and Giraitis, Kokoszka and Leipus (2000). In this case

$$\mathrm{var}(x_t) = \frac{\alpha_0}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i)}. \tag{3}$$

The following condition is required for identification of the GARCH parameters.

$$A(x) = \Sigma_{i=1}^{p} \alpha_i x^i \text{ and } B(x) = 1 - \Sigma_{i=1}^{q} \beta_i x^i \text{ are coprimes.} \tag{4}$$

Hall and Yao (2003) show that an explicit form for σ_t^2 is

$$\sigma_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \beta_i} + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} x_{t-i-j_1-\cdots-j_k}^2. \tag{5}$$

Put $y_t = \log(x_t^2)$ and $w_t = \log(z_t^2)$. Then we have

$$y_t = w_t + \log \sigma_t^2.$$

If the density f of z_t is symmetric around 0, then the density of w_t is g given by

$$g(w) = f(e^{w/2})e^{w/2}. \tag{6}$$

In particular when f corresponds to the N(0,1) distribution, $g = g_0$ where

$$g_0(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(e^w - w)}. \tag{7}$$

Given the parameter values $\mathbf{c} = (\mathbf{a}, \mathbf{b})$ where $\mathbf{a} = (a_0, a_1, ..., a_p)$, $\mathbf{b} = (b_1, ..., b_q)$ we define for all t

$$h_t(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{q} b_i h_{t-i}(\mathbf{c}). \tag{8}$$

where $x_t = 0$ for $t \leq 0$ and so

$$h_t(\mathbf{c}) = \frac{a_0}{1 - \sum_{i=1}^{q} b_i}$$

for all $t \leq 0$. From (5) we obtain that

$$\begin{aligned} h_t(\mathbf{c}) &= \frac{a_0}{1-\sum_{i=1}^q b_i} + \sum_{i=1}^p a_i x_{t-i}^2 \\ &+ \sum_{i=1}^p a_i \sum_{k=1}^\infty \sum_{j_i=1}^q \cdots \sum_{j_k=1}^q b_{j_1} \cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k} \, I_{t-i-j_1-\cdots j_k \geq 1}. \end{aligned}$$

These initial conditions are the same as those used by Hall and Yao (2003).

The usual form of the QML-estimate based on the x_t's consists on maximizing

$$-\frac{1}{2}\sum_{t=p+1}^T \frac{x_t^2}{h_t(\mathbf{c})} - \frac{1}{2}\sum_{t=p+1}^T \log h_t(\mathbf{c}) \tag{9}$$

and since $y_t = \log(x_t^2)$, this can be written as

$$-\frac{1}{2}\sum_{t=p+1}^T \left(e^{y_t - \log h_t(\mathbf{c})} + \log h_t(\mathbf{c})\right).$$

Then, to maximize (9) is equivalent to maximize

$$L_{0,T}(\mathbf{c}) = \sum_{t=p+1}^T \log(g_0(y_t - \log h_t(\mathbf{c})). \tag{10}$$

where the function g_0 is given by (7).

Observe that this equivalence does not require that the true density f of z_t be symmetric.

Maximizing (10) is equivalent to minimizing

$$M_{0,T}(\mathbf{c}) = \frac{1}{T-p}\sum_{t=p+1}^T \rho_0(y_t - \log h_t(\mathbf{c})), \tag{11}$$

where

$$\rho_0 = -\log(g_0), \tag{12}$$

where g_0 is given in (7).

In a similar way, it can be proved that the ML-estimate for a GARCH(p,q) model corresponding to any symmetric density f^* (it is not necessary that f^* coincides with the true density f) is obtained by minimizing

$$\frac{1}{T-p}\sum_{t=p+1}^T \rho^*(y_t - \log h_t(\mathbf{c})),$$

where $\rho^* = -\log(g^*)$ and g^* is given by (6) with $f = f^*$.

As could be expected, the QML-estimate is not robust, i.e., a few outliers may have a large influence on this estimate. This can be seen in our Monte Carlo simulation in Section 4. One reason for the lack of robustness of the QML-estimate is that ρ_0 is unbounded, so that one large outlier may have an unbounded effect on $M_{0,T}$.

Put

$$M_T(\mathbf{c}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log h_t(\mathbf{c})). \tag{13}$$

Then, the M-estimates for GARCH models are defined as

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\mathbf{c} \in C} M_T(\mathbf{c}) \tag{14}$$

for some convenient compact set C. These estimates can be considered a generalization of the class of M-estimates proposed by Huber (1964) for location and Huber (1973) for regression.

Define

$$J(u) = E(\rho(w_t - u)).$$

In Lemma 1 we show that $J(u)$ is well defined when ρ' is finite.

Lemma 1. *Consider a stationary GARCH model x_t given by (1). Then (a) $E(|w_t|) < \infty$ and (b) If $\psi = \rho'$ is finite, then $J(u)$ is finite for all u.*

To guarantee good consistency properties of the estimates we need that ρ satisfy the following property

P1. There exists a unique value u_0 where $J(u)$ takes the minimum.

Bollerslev, Chou and Kroner (1992) proposed using the ML-estimate for z_t having a symmetric heavy tail distribution, for example a Student distribution with a small degree of freedom. This corresponds to an M-estimate with $\rho = -\log(g)$, where g is the density of $\log(z^2)$. Peng and Yao (2003) LAD estimate corresponds to $\rho(u) = |u|$.

We can distinguish two types of M-estimates: (i) M-estimates with ρ' bounded but ρ unbounded (ii) M-estimates with both ρ and ρ' bounded. The M- estimates with ρ' bounded but ρ unbounded are robust when z_t has heavy tail distribution, although they may be much affected by another type of outliers as for example additive outliers, as we see in our Monte Carlo simulation in Section 4. To increase the degree of robustness we need that ρ be bounded too. There is extensive literature on the properties of M-estimates for regression. For example, Huber (1973) shows that M-estimates for regression with bounded ρ' are robust when the distribution of the error is heavy tailed. Yohai (1987) shows that M-estimates for regression with ρ bounded are robust against any kind of outliers.

The ML-estimates for heavy tail z_t and the Peng and Yao (2003) LAD estimates are examples of M-estimates with bounded ρ' but unbounded ρ. For instance for the Student distribution with three degrees of freedom we have

$$\rho(u) = \quad 2\ln\left(1+e^{u}\right) - u/2$$

and

$$\rho'(u) = \frac{3e^{u}-1}{2(1+e^{u})}.$$

For the Peng and Yao estimate we have

$$\rho(u) = |u|, \rho'(u) = \operatorname{sign}(u).$$

M-estimates with ρ bounded are more robust than the QML-estimate, although large outliers may still have a strong effect on the estimates. The reason is that this estimate requires computing the values $h_t(\mathbf{c})$ using (8), so a large outlier at time t may affect all the $h_{t'}(\mathbf{c})$ with $t' > t$.

The same problem appears in the estimates of ARMA models, where an outlier at time t may influence the estimated innovations corresponding to several periods. To deal with this problem, several authors used robust filters. See Denby and Martin (1979), Martin, Samarov and Vandaele (1983) and Bianco, García Ben, Martínez and Yohai (1996). Muler and Yohai (2002) used robust filters for estimating ARCH models. However, the asymptotic theory of these estimates is very complicated and proofs of asymptotic normality are not available.

In this paper we propose a method related to robust filters, which has the advantage that the resulting estimates are mathematically tractable. To gain robustness, we modify the M-estimates for GARCH models by including a mechanism which restricts the propagation of the outlier effect on the estimated $h_t(\mathbf{c})$'s. For this purpose, we replace in the computation of the M-estimate $h_t(\mathbf{c})$ by

$$h^*_{t,k}(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i h^*_{t-i,k}(\mathbf{c}) r_k\left(\frac{x^2_{t-i}}{h^*_{t-i,k}(\mathbf{c})}\right) + \sum_{i=1}^{q} b_i h^*_{t-i,k}(\mathbf{c}), \tag{15}$$

where $x_t = 0$ for $t \leq 0$ and where

$$r_k(u) = \begin{cases} u & \text{if} \quad u \leq k \\ k & \text{if} \quad u > k. \end{cases} . \tag{16}$$

Observe that if k is large, then $h_t(\mathbf{c})$ and $h^*_{t,k}(\mathbf{c})$ are close. However the propagation of the effect of one outlier in time t on the $h^*_{t',k}(\mathbf{c})$, $t' > t$, practically vanishes after a few periods. Therefore, if x_t follows a GARCH model but contains some outliers, we may expect that the M-estimates using conditional variances given by (15) would fit better than the M-estimate corresponding to the GARCH model. This suggests modifying the M-estimate as follows. Let $\widehat{\boldsymbol{\gamma}}_1$ be defined as in (14) and $\widehat{\boldsymbol{\gamma}}_2$ by

$$\widehat{\boldsymbol{\gamma}}_2 = \arg\min_{\mathbf{c}\in C} M^*_{Tk}(\mathbf{c}), \tag{17}$$

where

$$M^*_{Tk}(\mathbf{c}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log h^*_{t,k}(\mathbf{c})). \tag{18}$$

When the process is a perfectly observed GARCH process without outliers the conditional variances are given by (8). Then the estimate $\widehat{\boldsymbol{\gamma}}_1$using these conditional variances generally behaves better than $\widehat{\boldsymbol{\gamma}}_2$. In this case $\widehat{\boldsymbol{\gamma}}_2$ is asymptotically biased and we may expect $M_T(\widehat{\boldsymbol{\gamma}}_1) \leq M^*_{Tk}(\widehat{\boldsymbol{\gamma}}_2)$. Theorem 4 of Section 3 proves that this holds asymptotically with probability one. As explained above, when there are outliers, $\widehat{\boldsymbol{\gamma}}_2$ may be preferable and we may expect $M_T(\widehat{\boldsymbol{\gamma}}_1) > M^*_{Tk}(\widehat{\boldsymbol{\gamma}}_2)$. Then we define the BM-estimate by

$$\boldsymbol{\gamma}^B = \begin{cases} \widehat{\boldsymbol{\gamma}}_1 & \text{if} \quad M_T(\widehat{\boldsymbol{\gamma}}_1) \leq M^*_{Tk}(\widehat{\boldsymbol{\gamma}}_2) \\ \widehat{\boldsymbol{\gamma}}_2 & \text{if} \quad M_T(\widehat{\boldsymbol{\gamma}}_1) > M^*_{Tk}(\widehat{\boldsymbol{\gamma}}_2). \end{cases} \tag{19}$$

We will see that BM-estimates simultaneously possess both properties: robustness against outliers and consistency when the series follows a GARCH model without outliers. Moreover, by choosing m and k conveniently, these estimates have high efficiency under the GARCH model

Our proposal is to use BM-estimates with ρ of the form $\rho = m(\rho_0)$, where m is a bounded nondecreasing function. We see in the next section that this function satisfies P1 with $u_0 = 0$ when z_t is normal. Moreover, as shown in Section 4, when we take m equal to the identity in a sufficiently large interval, the BM-estimates are going to be highly efficient when z_t has normal distribution and less sensitive to additive outliers than the other estimates mentioned in this section.

However, if we consider that z_t has a symmetric density f different from the normal, it is possible to define $\rho = m(-\log(g))$ where g is given by (6) and ρ is a non decreasing and bounded function.

3 Asymptotic results

In this Section we state the main asymptotic results for the M- and BM-estimates: consistency and asymptotic normality. In Theorems 1, 2 and 3 we prove the consistency and asymptotic normality for any M-estimate defined in (14) as long as P1 holds and ρ' is bounded. In Theorem 4 and 5 we prove the consistency and asymptotic normality for the proposed BM-estimators.

Suppose first that we have the infinite sequence of observations $\mathbf{X}_t = (..., x_{t-1}, x_t)$ corresponding to a GARCH(p, q) process up to time t with parameter $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, and given $\mathbf{c} = (\mathbf{a}, \mathbf{b})$ call $\widetilde{h}_t(\mathbf{c})$ the conditional variance of x_t given $\mathbf{X}_{t-1}$ when $\boldsymbol{\gamma} = \mathbf{c}$. Then the following recursive relationship is satisfied

$$\widetilde{h}_t(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{q} b_i \widetilde{h}_{t-i}(\mathbf{c}). \tag{20}$$

The following Theorem shows the Fisher consistency of the M-estimates of the GARCH model and gives a sufficient condition for P1.

Denote by

$$R_+^n = \{\mathbf{x} = (x_1, ...x_n) : \ x_i \geq 0, 1 \leq i \leq n\}.$$

Theorem 1. *Let x_t be a stationary GARCH(p, q) process satisfying (1) and (2). Let $y_t = \log(x_t^2)$ and define for $\mathbf{c} = (\mathbf{a}, \mathbf{b}) \in R_+^{p+q+1}$*

$$M(\mathbf{c}) = E(\rho(y_t - \widetilde{h}_t(\mathbf{c})).$$

Suppose that ρ' is bounded, that P1 and (4) hold and that $\beta_q > 0$ in the case of a GARCH(p,q) process or $\alpha_p > 0$ in the case of an ARCH(p) process. Then

(i) *$M(\mathbf{c})$ is minimized when $a_i = e^{u_0}\alpha_i$, $0 \leq i \leq p$, $b_i = \beta_i$, $1 \leq i \leq q$.*

(ii) *Assume that $w_t = \log(z_t^2)$ has a density $g(w)$ that is unimodal, continuous and positive for all w. If we take $\rho = m(-\log(g))$, where m is monotone, P1 holds with $u_0 = 0$.*

Observe that according to part (i) of Theorem 1, the M-estimate of $\boldsymbol{\alpha}$ should be corrected by the factor e^{-u_0} for consistency. Part (ii) shows that if we take $\rho = m(-\log(g))$ there is no need of correction. An alternative that avoids the correction factor is to replace $\rho(u)$ with $\overline{\rho}(u) = \rho(u - u_0)$, and then in the rest of the paper without loss of generality we assume $u_0 = 0$.

Put

$$\begin{aligned} C_\delta &= \Bigg\{ (\mathbf{a}, \mathbf{b}) : \mathbf{a} \in R_+^{p+1},\ \mathbf{b} \in R_+^q,\ \ a_0 \in [\delta, 1/\delta],\ \ \sum_{i=1}^{p} a_i \geq \delta\ , \\ & \sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j \leq 1 - \delta \Bigg\} \end{aligned} \tag{21}$$

The set C in (14) and (17) is taken as C_{δ_0} for some $\delta_0 > 0$.

Put $\boldsymbol{\gamma} = (\alpha_0, \alpha_1, ...\alpha_p, \beta_1, ..., \beta_q)$. The following two Theorems state the consistency and asymptotic normality of the M-estimates to $\boldsymbol{\gamma}$.

Theorem 2. *Suppose that all the assumptions of Theorem 1 hold. Let $\widehat{\boldsymbol{\gamma}}_T$ be defined as in (14) with $C = C_{\delta_0}$ given by (21). We also assume that P1 is satisfied with $u_0 = 0$, that ρ has a bounded derivative, and that $\boldsymbol{\gamma} \in C$. Then $\widehat{\boldsymbol{\gamma}}_T \to \boldsymbol{\gamma}$ a.s..*

Theorem 3. *Suppose that all the assumptions of Theorem 2 hold. Assume also that (i) ρ has three continuous and bounded derivatives, (ii) $E(\psi^2(w_t)) > 0$ and (iii) $E(\psi'(w_t)) > 0$, where $\psi = \rho'$. Then $T^{1/2}(\widehat{\boldsymbol{\gamma}}_T - \boldsymbol{\gamma})$ converges in distribution to a $N(\mathbf{0}, V)$ and*

$$V = \frac{E_g(\psi^2(w_t))}{E_g^2(\psi'(w_t))} \left(E_g \left(\frac{1}{\widetilde{h}_t^2(\boldsymbol{\gamma})} \nabla \widetilde{h}_t(\boldsymbol{\gamma}) \nabla \widetilde{h}_t(\boldsymbol{\gamma})' \right) \right)^{-1}, \tag{22}$$

where ∇h denotes gradient of h.

For the case of the QML-estimate, we have $\rho = \rho_0$ given in (12). Then the assumption of ρ' bounded is not satisfied. However, in the case that z_t has a finite fourth moment, the QML-estimate has asymptotic normal distribution with a covariance matrix given by (22). See Berkes, Hováth and Kokoszka (2003). Then the relative asymptotic efficiency of the M-estimate with respect to the QML-estimate is given by

$$AEF = \frac{a(\psi, g)}{a(\psi_0, g)},$$

where $\psi_0 = \rho_0'$ and

$$a(\psi, g) = \frac{E_g(\psi^2(w_t))}{E_g^2(\psi'(w_t))}.$$

Therefore, by choosing m bounded and close to the identity function, we can obtain a robust estimate that is highly efficient when the z_t's are normal.

The next two Theorems show that asymptotically the M and BM-estimates are equivalent when x_t follows an exact GARCH model without outliers.

Theorem 4. *Suppose that all the assumptions of Theorem 3 hold and that the distribution of z_t gives positive probability to the complement of any compact. We also assume that* $\lim_{|u|\to\infty} \rho(u) = \sup_u \rho(u)$. *Moreover if we let $\widehat{\boldsymbol{\gamma}}_T$ be the M-estimate defined by (14) and $\boldsymbol{\gamma}_T^B$ the BM-estimate defined by (19), then* $\lim_{T\to\infty} P(\boldsymbol{\gamma}_T^B = \widehat{\boldsymbol{\gamma}}_T) = 1$.

Remark. Suppose that g is a unimodal and positive density. Then, it can be proved that the assumption $\lim_{|u|\to\infty} \rho(u) = \sup_u \rho(u)$ holds if we take $\rho = m(-\log(g))$ and m non decreasing.

From Theorems 2, 3 and 4 we derive the following result.

Theorem 5. *Theorems 2 and 3 also hold for the BM-estimate $\boldsymbol{\gamma}_T^B$.*

4 Monte Carlo Simulation

We performed a Monte Carlo study to compare the behavior of seven estimates: (i) the QML-estimate (QML), (ii) the maximum likelihood corresponding to z_t with Student distribution with three degrees of freedom (SML) (iii) the LAD Peng-Yao estimate (LAD), (iv) the M-estimate based on a loss function $\rho_1 = m_1(\rho_0)$, where ρ_0 is given in (12) and m_1 is a nondecreasing, bounded and close to the identity function which is defined below (M_1), (v) A BM-estimate as defined in (19) with $\rho = \rho_1$ and $k = 5.02$ (BM_1), (vi) an M-estimate based on a loss function ρ_2 defined as $\rho_2(x) = m_2(\rho_0(x))$, $m_2(v) = 0.8m_1(v/0.8)$ (M_2) and (vii) A BM-estimate as defined in (19) with $\rho = \rho_2$ and $k = 3$ (BM_2).

The function m_1 is defined as

$$m_1(x) = \begin{cases} x & \text{if} \quad x \leq 4.02 \\ c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0 & \text{if} \quad 4.02 < x \leq 4.30 \\ 4.16 & \text{if} \quad x > 4.30, \end{cases}$$

where $c_0 = 6777, c_1 = -6536.2, c_2 = 2362.3,\ c_3 = -379.0087,\ c_4 = 22.7770.$ This function is shown in Figure 1.

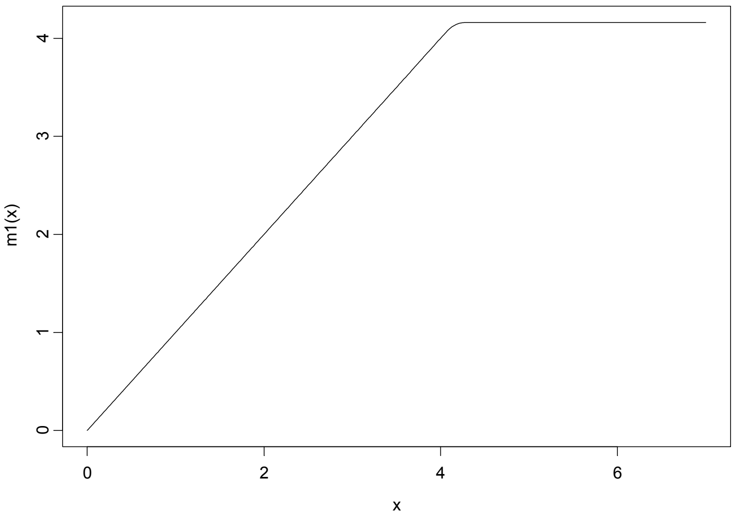

Figure 1: Function m_1

As we see in Figure 1, m_1 is a smoothed version of

$$m(x) = \begin{cases} x & \text{if} \quad x \leq 4.02 \\ 4.02 & \text{if} \quad x > 4.02. \end{cases}$$

The function m_1 is equal to the identity in a larger interval than m_2 and therefore the estimates based on m_1 are more similar to the QML than those based on m_2. These, (see in Table 1), make estimates M_1 and BM_1 more efficient than estimates M_2 and BM_2. As a counterpart we see in our Monte Carlo results that the first estimates are going to be less robust than the second ones. The choice of m_1 and m_2 were done so

$$P(\rho_1(w_t) = \rho_0(w_t)) = 0.96. \tag{23}$$

and

$$P(\rho_2(w_t) = \rho_0(w_t)) = 0.90$$

when w_t is $\log(z_t^2)$ and z_t is N(0,1).

After several trials, the value of k in (16) for BM_1 was taken as equal to 5.02. This value is such $P(z_t^2 \leq 5.02) = 0.975$, when z_t is N(0,1). For the BM_2 estimate, in order to gain robustness, we chose $k = 2.72$. This value is such $P(z_t^2 \leq 2.72) = 0.90$. We found in the Monte Carlo simulations that BM_1 was a convenient trade off between efficiency under a normal GARCH model and robustness. Instead, BM_2 has rather low efficiency under a GARCH model, but we found in the Monte Carlo simulation that is more robust when the fraction of outliers is 10%.

The correction term u_0 defined in P1 when z_t is N(0,1) is 0.636 for the SML estimate and -0.787 for the LAD estimate. For the other estimates it is zero. The asymptotic efficiencies (EFF) of all the estimates we used in these simulations under a normal GARCH model are shown in Table 1. We observe that the asymptotic efficiencies of the M_1, BM_1 and SML estimates are quite high, the asymptotic efficiencies of M_2 and BM_2 are intermediate and the asymptotic efficiency of the LAD estimate is quite low.Table 1. Asymptotic efficiencies (EFF) of the estimates under normal GARCH models.

Table 1. Asymptotic efficiencies (EFF) of the estimates under normal GARCH models.

Estimate	QML	SML	LAD	M_1 and BM_1	M_2 and BM_2
EFF	1	0.79	0.37	0.83	0.67

We report the results using a GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and an ARCH(2) model with parameters $\alpha_0 = 1$,$\alpha_1 = 0.5$ and $\alpha_2 = 0.4$. Other GARCH(1,1) and ARCH(2) models were simulated, and the results were similar to those mentioned above. In all cases the number of observations n was 1000 and the number of Monte Carlo replications was 500. The constant δ_0 used to define the compact set C in (21) was taken as equal to 0.01.

For each model we consider four cases: (a) z_t normal and no outliers (b) z_t normal with 5% of additive outliers and, (c) z_t normal with 10% of additive outliers and (d) z_t has a Student distribution with 3 degrees of freedom.

The series x_t^* with additive outliers is defined as follows

$$x_t^* = \begin{cases} x_t + d\sigma_t & \text{if} \quad t = t_i,\ 1 \leq i \leq l = hn/100 \\ x_t & \qquad \text{elsewhere,} \end{cases}$$

where h is the percentage of contamination, x_t is the non contaminated series in GARCH models with z_t normal, $t_1, ..., t_l$ are the times when the outliers are observed. The values $t_i, 1 \leq i \leq l$, were chosen equally spaced. We considered two values for d : 3 and 5.

Tables 2 and 3 show the mean square errors (MSE) in the case of no outliers for the normal GARCH(1,1) model and for the normal ARCH(2) model respectively. In both tables we show the efficiency (EFF) of the estimates for these finite samples with respect to the QML. We observe that in this case M_1 and BM_1 behave similarly. The same happens with M_2 and BM_2.

Table 2. Mean square errors (MSE) and efficiencies with respect to the QML (EFF) for a normal GARCH(1,1) model without outliers with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$.

Estimate	α_0		α_1		β_1	
	MSE	EFF	MSE	EFF	MSE	EFF
QML	0.033	1.00	0.004	1.00	0.003	1.00
SML	0.042	0.80	0.005	0.79	0.003	0.80
LAD	0.092	0.36	0.011	0. 33	0.008	0.35
M_1	0.040	0.84	0.004	0.90	0.003	0.87
BM_1	0.040	0.85	0.004	0.88	0.003	0.87
M_2	0.068	0.49	0.007	0.51	0.004	0.72
BM_2	0.067	0.50	0.008	0.45	0.004	0.70

Table 3. Mean square errors (MSE) and efficiencies with respect to the QML (EFF) for a normal ARCH(2) model without outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\alpha_2 = 0.4$.

Estimate	α_0		α_1		α_2	
	MSE	EFF	MSE	EFF	MSE	EFF
QML	0.012	1.00	0.004	1.00	0.0035	1.00
SML	0.015	0.80	0.005	0. 80	0.0042	0.83
LAD	0.032	0.38	0.012	0.33	0.0093	0.38
M_1	0.014	0.86	0.005	0.87	0.0040	0.88
BM_1	0.014	0.87	0.005	0.87	0.0041	0.85
M_2	0.028	0.43	0.009	0.44	0.0067	0.52
BM_2	0.026	0.45	0.009	0.44	0.0067	0.52

In Tables 4 and 5 we show the MSE for 5% contaminated samples for the normal GARCH(1,1) model and for the normal ARCH(2) model respectively. We observe in these tables that QML can be seriously affected by outliers, especially for $d = 5$. Although LAD, SML, M_1 and M_2 are not so much affected by outliers we can see that both, BM_1 and BM_2, in general, behave much better.

Table 4. Mean square errors for a normal GARCH(1,1) model with 5% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$ Outlier size: $d\sigma_t$.

Estimate	$d = 3$			$d = 5$		
	α_0	α_1	β_1	α_0	α_1	β_1
QML	2.11	0.037	0.021	23.27	0.38	0.104
SML	1.13	0.015	0.017	5.82	0.05	0.088
LAD	0.83	0.046	0.022	2.22	0.13	0.065
M_1	1.38	0.022	0.040	0.49	0.05	0.029
BM_1	0.39	0.011	0.012	0.07	0.01	0.006
M_2	1.25	0.022	0.032	0.76	0.051	0.040
BM_2	0.34	0.013	0.010	0.09	0.007	0.006

Table 5. Mean square errors for a normal ARCH(2) model with 5% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\alpha_2 = 0.4$ Outlier size: $d\sigma_t$.

Estimate	$d = 3$			$d = 5$		
	α_0	α_1	α_2	α_0	α_1	α_2
QML	0.61	0.025	0.017	4.79	0.26	0.1613
SML	0.35	0.014	0.014	1.23	0.04	0.0558
LAD	0.37	0.040	0.034	1.06	0.11	0.0865
M_1	0.26	0.022	0.019	0.08	0.03	0.0338
BM_1	0.12	0.012	0.012	0.02	0.010	0.0119
M_2	0.29	0.020	0.018	0.18	0.033	0.0348
BM_2	0.11	0.014	0.010	0.04	0.006	0.0073

In Tables 6 and 7 we report the MSE for the normal GARCH(1,1) and normal ARCH(2) models respectively when there is 10% outlier contamination. In the case of $d = 5$ the behavior of the estimates is similar to the case with 5% of outliers. In the case of $d = 3$ the only estimate that is not much affected by outliers is BM_2.

Table 6 Mean square errors for a normal GARCH(1,1) model with 10% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$ Outlier size: $d\sigma_t$.

Estimate	$d = 3$			$d = 5$		
	α_0	α_1	β_1	α_0	α_1	β_1
QML	15.80	0.06	0.07	95.83	0.41	0.19
SML	7.24	0.06	0.06	9.87	0.22	0.17
LAD	2.58	0.13	0.05	2.44	0.23	0.12
M_1	12.76	0.12	0.20	0.23	0.14	0.048
BM_1	6.22	0.04	0.10	0.07	0.03	0.007
M_2	8.52	0.12	0.13	0.21	0.14	0.055
BM_2	1.58	0.02	0.03	0.07	0.01	0.006

Table 7. Mean square errors for a normal ARCH(2) model with 10% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\alpha_2 = 0.4$ Outlier size: $d\sigma_t$.

Estimate	$d = 3$			$d = 5$		
	α_0	α_1	α_2	α_0	α_1	α_2
QML	3.96	0.047	0.030	44.50	0.66	0.49
SML	2.82	0.051	0.052	12.72	0.16	0.12
LAD	1.92	0.106	0.084	4.67	0.21	0.14
M_1	5.65	0.195	0.177	0.56	0.09	0.098
BM_1	2.18	0.049	0.067	0.03	0.02	0.026
M_2	3.80	0.127	0.118	0.71	0.099	0.094
BM_2	0.68	0.020	0.025	0.04	0.009	0.013

Table 8 reports the MSE for the Student GARCH(1,1). As may be expected the smallest MSE corresponds to the SML. The other robust estimates behave quite similarly and better than the QML.

Table 8. Mean square errors for a student GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$.

Estimate	α_0	α_1	β_1
QML	0.275	0.109	0.023
SML	0.048	0.011	0.007
LAD	0.080	0.019	0.011
M_1	0.067	0.018	0.010
BM_1	0.070	0.018	0.010
M_2	0.089	0.022	0.013
BM_2	0.090	0.023	0.013

In Figures 2-4 we plot the MSE's as a function of the outlier size d for QML, SML, LAD, BM_1 and BM_2 for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination. We observed that both BM_1 and BM_2 behave more robustly than the others.

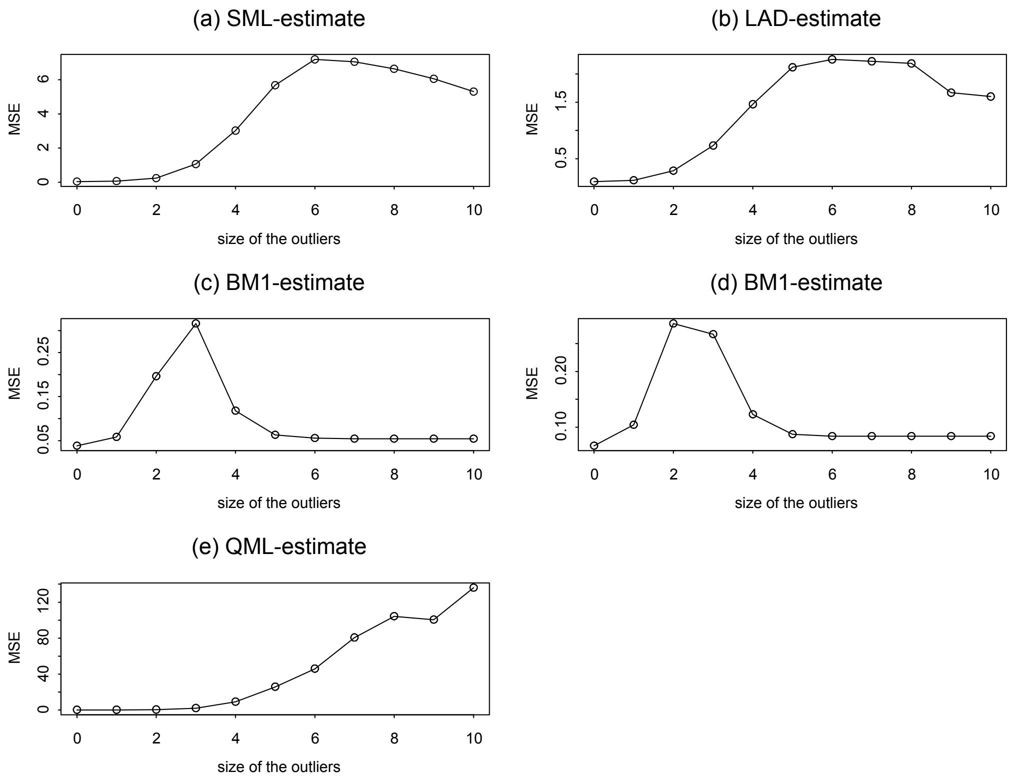

Figure 2: Mean Square Errors of α_0 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination.

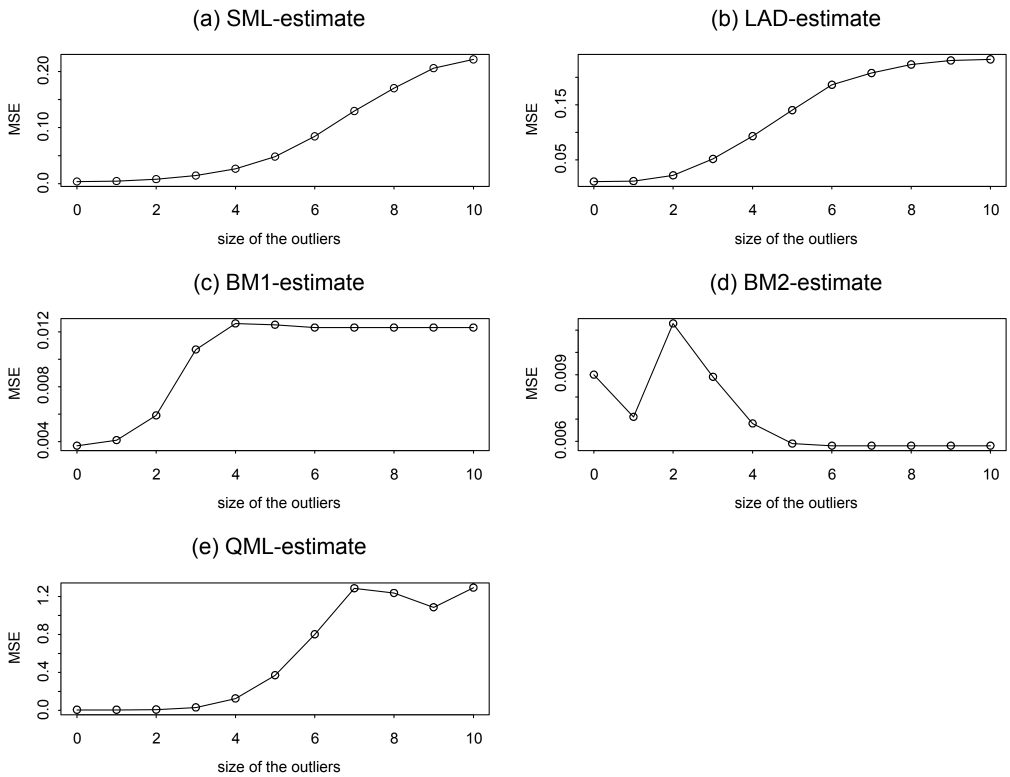

Figure 3: Mean Square Errors of α_1 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination.

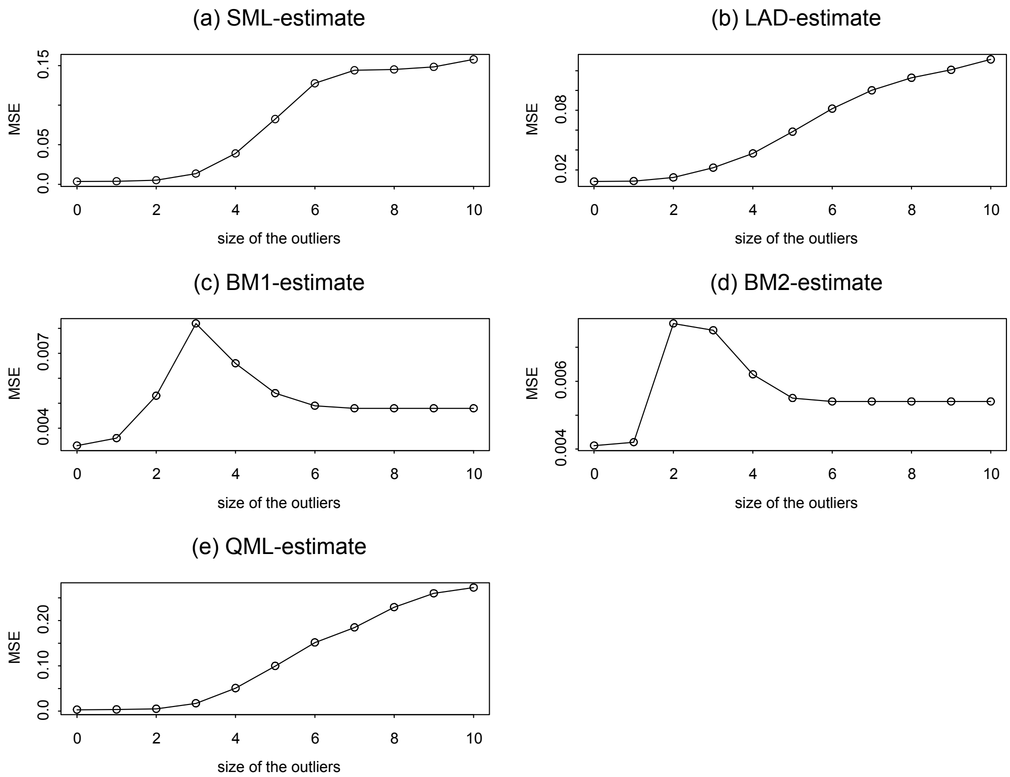

Figure 4: Mean Square Errors of β_1 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination.

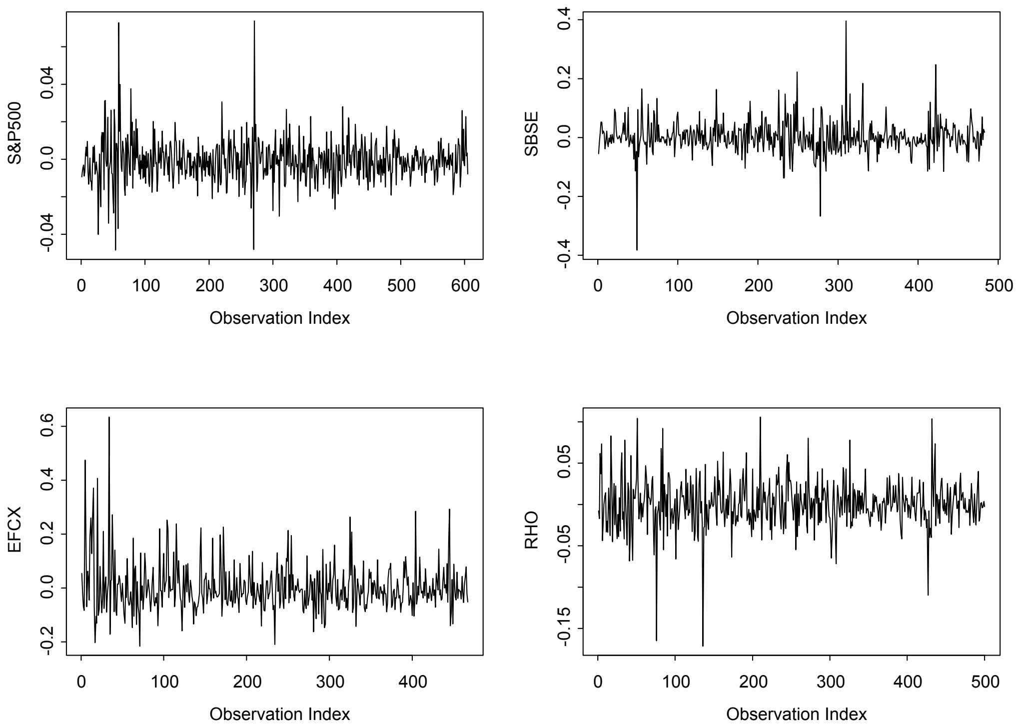

Figure 5: Plot of the Daily Return Series.

5 Analysis of Some Examples

We consider four different examples of series corresponding to daily financial data: (a) The Standard and Poor 500 Index (S&P 500) from February 1, 2000 to June 30, 2002 (b) The SBS Technologies Inc.(SBSE) from January 3, 2000 to December 31, 2001 (c) Electric Fuel. Corp (EFCX) from January 3, 2000 to December 31, 2001 and (d) Rohm and Haas Company (ROH) from January 3, 2000 to December 31, 2001. In Fig 5 we plot the daily returns of these four series. We observe that the series contain several outliers that correspond to unusually large movements in the prices.

After centering with the median, each of these series was fitted as a

GARCH(1,1) model using the QML-estimate, SML-estimate, LAD-estimate and BM$_1$-estimate defined as in the Section 4. In Table 9 we show these estimates. Since the series contain outliers, as can be expected, the estimates show some important differences.

Table 9. Fitted GARCH(1,1) models for the daily returns series.

	S&P 500			SBSE		
Estimates	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$
QML	$3.9 * 10^{-6}$	0.11	0.86	$3.2 * 10^{-4}$	0.28	0.68
SML	$5.1 * 10^{-6}$	0.08	0.85	$2.1 * 10^{-4}$	0.09	0.76
LAD	$4.9 * 10^{-6}$	0.06	0.85	$4.5 * 10^{-4}$	0.07	0.56
BME$_1$	$4.5 * 10^{-6}$	0.10	0.84	$4.1 * 10^{-4}$	0.26	0.50
	EFCX			RHO		
Estimates	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$
QML	$3.0 * 10^{-4}$	0.050	0.91	$3.1 * 10^{-5}$	0.054	0.91
SML	$1.0 * 10^{-3}$	0.106	0.65	$3.4 * 10^{-5}$	0.048	0.88
LAD	$1.02 * 10^{-3}$	0.138	0.54	$9.2 * 10^{-5}$	0.076	0.71
BME$_1$	$2.6 * 10^{-3}$	0.259	0.20	$2.4 * 10^{-4}$	0.306	0.30

Let x_t, $1 \leq t \leq T$, be an observed centered series and $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ an estimate for the GARCH(1,1) model. Let $\widehat{\sigma}_t^2$ be the conditional variance of x_t obtained using the estimated parameters. In the case of the QML-, SML- and LAD-estimates, $\widehat{\sigma}_t^2$ is recursively computed by

$$\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 x_{t-1}^2 + \widehat{\beta}_1 \widehat{\sigma}_{t-1}^2, \ 2 \leq t \leq T. \tag{24}$$

In the case of the BM$_1$- and the BM$_2$-estimate $\widehat{\sigma}_t^2$ is also given by (24) if $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ coincides with the corresponding $M-$estimate $\widehat{\boldsymbol{\gamma}}_1$ defined in (14) or by

$$\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{\sigma}_{t-1}^2 g_k(x_{t-1}^2/\widehat{\sigma}_{t-1}^2) + \widehat{\beta}_1 \widehat{\sigma}_{t-1}^2, \ 2 \leq t \leq T$$

if $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ coincides with the bounded estimate $\widehat{\boldsymbol{\gamma}}_2$ defined in (17).

When x_t follows a GARCH model, the series z_t have the following two properties (a) var(z_t)=1 and (b) z_t^2 is uncorrelated to z_{t-1}^2. We use these properties to evaluate the performance in the four data sets of the different estimates.

Given an estimate $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ let us define

$$\widehat{z}_t = \frac{x_t}{\widehat{\sigma}_t} \quad , 2 \leq t \leq T. \tag{25}$$

If this estimate use to define $\widehat{\sigma}_t$ is close to the true value, properties (a) and (b) should approximately hold for the $\widehat{z}_t'$s.

Since the sample variance is not robust, to compare how property (a) is satisfied for the different estimates we use a normalized 0.10-trimmed sample variance of $\widehat{z}_t$ defined by

$$\sigma^2_{\mathrm{TR}} = \frac{1.605}{T_1} \sum_{t=1}^{T_1} \widehat{z}^2_{(t)}, \tag{26}$$

where $\widehat{z}^2_{(1)} \leq ... \leq \widehat{z}^2_{(T-1)}$ are the order statistics of $(\widehat{z}^2_2, ..., \widehat{z}^2_T)$, T_1 is the integer part of $0.9(T-1)$. The value 1.605 was chosen so that the normalized trimmed variance be one for normal samples. To compare the estimates in reference to property (b) we compute the rank correlation between $\widehat{z}^2_{t-1}$ and $\widehat{z}^2_t$, which is a robust correlation measure. We denote this estimate by τ.

Table 10 shows the value of σ^2_{TR} and τ corresponding to the QML-, SML-, LAD- and BM$_1$-estimates for the four series.

Table 10. Truncated variance (σ^2_{TR}) and rank correlation (τ) for the daily returns series.

	S&P 500		SBSE		EFCX		RHO	
Estimates	σ^2_{TR}	τ	σ^2_{TR}	τ	σ^2_{TR}	τ	σ^2_{TR}	τ
QML	0.79	0.043	0.58	0.044	0.60	0.083	0.685	0.060
SML	0.93	0.045	0.98	0.093	0.97	0.025	0.928	0.059
LAD	1.07	0.056	1.24	0.072	1.18	−0.009	1.144	0.041
BME$_1$	0.98	0.022	1.02	−0.015	0.96	−0.043	1.045	−0.035

We observe that value of σ^2_{TR} for QML is in general much lower than one. Instead the robust estimates give values closer to one. In general all the estimates give values of τ close to zero. Taking into account both indicators σ^2_{TR} and τ, the BM-estimate performs better than the others for series S&P 500, SBSE and RHO. However, for EFCX the SML seems preferable to the others.

6 Concluding Remarks.

In this paper we present two classes of robust estimates for GARCH models: M- and BM-estimates. A Monte Carlo study shows that for the GARCH(1,1) model, the QML-estimate may practically collapse when there is 5% outlier contamination. All the robust estimates are in general less influenced by outliers. However, the BM-estimates generally behave much better than the rest of the robust methods under outlier contamination.

The study of four examples of daily returns series containing outliers shows that all the robust estimates are better than the QML-estimate. The BM-estimate seems to behave better than the others in three of these examples.

Our proposal is to always compute the BM- and the QML-estimates when fitting a GARCH model. A strong discrepancy between the two estimates indicates the presence of outliers in the series. In this case the decision of which estimate is preferable can be based on the comparison of the statistics σ^2_{TR} and τ for both fits. Of course this strategy can include other robust estimates such as the LAD or the SML.

7 Appendix

Proof of Lemma 1.

We have that

$$E(|w_t|) = \int_{-\infty}^{\infty} \left|\log(z^2)\right| f(z)dz. \tag{27}$$

Since f is continuous, there is a constant K such that $|f(z)| \leq K$ for all $z \in [-1, 1]$. Then we have

$$\int_{|z|\leq 1} \left|\log(z^2)\right| f(z)dz \leq K \int_{|z|\leq 1} \left|\log(z^2)\right| dz < \infty. \tag{28}$$

Since for $u \geq 1$ it holds that $\log(u) < u$. Then using that z_t has finite second moment we get

$$\int_{|z|>1} \left|\log(z^2)\right| f(z)dz \leq \int_{|z|>1} z^2 f(z)dz < \infty. \tag{29}$$

Then part (a) of the Lemma follows from (27),(28) and (29).

Then (b) follows from (a) and the fact that ρ satisfies a global Lipschitz condition.

Proof of Theorem 1.

(i) Let $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ be the true parameter. Then, we can write

$$M(\mathbf{c}) = E\left(\rho\left(w_t - \log\left(\frac{\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\boldsymbol{\gamma})}\right)\right)\right), \tag{30}$$

where $w_t = \log(y_t/\widetilde{h}_t(\boldsymbol{\gamma})) = \log(z_t^2)$ are i.i.d. random variables with distribution g. Since $\widetilde{h}_t(\mathbf{c})$ depends only on x_{t^*} with $t^* < t$ and $E(\rho(w_t - u))$ has a unique minimum at u_0, the minimum of $M(\mathbf{c})$ is attained at a point $\overline{\mathbf{c}}$ such that

$$\widetilde{h}_t(\overline{\mathbf{c}}) = e^{u_0}\widetilde{h}_t(\boldsymbol{\gamma}) \text{ a.s..}$$

Let $\boldsymbol{\gamma}^* = (e^{u_0}\alpha_0, e^{u_0}\alpha_1, ..., e^{u_0}\alpha_p, \beta_1, ..., \beta_q)$, then we have $\widetilde{h}_t(\boldsymbol{\gamma}^*) = e^{u_0}\widetilde{h}_t(\boldsymbol{\gamma})$ and so $\widetilde{h}_t(\overline{\mathbf{c}}) = \widetilde{h}_t(\boldsymbol{\gamma}^*)$. Therefore, from Corollary 2.1 of Berkes, Horvath and Kokoszka (2003), we obtain $\overline{\mathbf{c}} = \boldsymbol{\gamma}^*$.

(ii) Since g is strictly unimodal, continuous and $g(u) > 0$ for all u, Lemma 1 of Bianco, Garcia Ben and Yohai (2005) implies that $E(\rho(w_t - u))$ has a unique minimum at $u_0 = 0$. Hence, (ii) follows

According to the Remark after Theorem 1, in the rest of the Appendix we will assume $u_0 = 0$ and $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}$ without loss of generality.

The following Lemmas 2, 3 and 4 are used in the proofs of Theorems 2 and 3.

Lemma 2. Let x_t be a stationary and ergodic GARCH(p, q) process satisfying (1) and(2). Let $h_t(\mathbf{c})$ be as defined in (8) and $\widetilde{h}_t(\mathbf{c})$ as defined in (20). Then

(i) There exists $0 < \vartheta < 1$ and a positive finite random variable W such that $\sup_{\mathbf{c}\in C}\left|\widetilde{h}_t(\mathbf{c}) - h_t(\mathbf{c})\right| \leq \vartheta^t W$ for all $t \geq p+1$.

(ii) *There exists a neighborhood U of $\boldsymbol{\gamma}$ such that $\sup_{\mathbf{c}\in U} E\left|\nabla\log(\widetilde{h}_t(\mathbf{c}))\right|^n < \infty$ for all n.*

(iii)There exists a neighborhood U of $\boldsymbol{\gamma}$, $0 < \vartheta < 1$ and a finite positive finite random variable W_1 such that

$$\sup_{\mathbf{c}\in U}\left|\nabla\log(\widetilde{h}_t(\mathbf{c})) - \nabla\log(h_t(\mathbf{c}))\right| \leq \vartheta^t W_1$$

for all $t \geq p+1$.

(iv)There exists a neighborhood U of $\boldsymbol{\gamma}$, $0 < \vartheta < 1$ and a positive finite random variable W_2 such that

$$\sup_{\mathbf{c}\in U}\left\|\nabla\log(\widetilde{h}_t(\mathbf{c}))\nabla\log(\widetilde{h}_t(\mathbf{c}))'-\nabla\log(h_t(\mathbf{c}))\nabla\log(h_t(\mathbf{c}))'\right\|\leq\vartheta^t W_2$$

for all $t\geq p+1$,*where* $||A||$ *denotes the* l_2 *norm of* A.

(v)There exists a neighborhood U *of* $\boldsymbol{\gamma}$ *such that* $E\left(\sup_{\mathbf{c}\in U}\left\|\nabla^2\widetilde{h}_t(\mathbf{c})\right\|^2\right)<\infty$.

(vi)There exists a neighborhood U *of* $\boldsymbol{\gamma}$, $0<\vartheta<1$ *and positive finite random variable* W_3 *such that*

$$\sup_{\mathbf{c}\in U}\left\|\frac{\nabla^2\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})}-\frac{\nabla^2 h_t(\mathbf{c})}{h_t(\mathbf{c})}\right\|\leq\vartheta^t W_3$$

for all $t\geq p+1$.

Proof of (i)

Hall and Yao (2003) show that

$$\widetilde{h}_t(\mathbf{c})=\frac{a_0}{1-\sum_{i=1}^q b_i}+\sum_{i=1}^p a_i x_{t-i}^2+\sum_{i=1}^p a_i\sum_{k=1}^\infty\sum_{j_i=1}^q\cdots\sum_{j_k=1}^q b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k}, \tag{31}$$

and then

$$\begin{aligned} h_t(\mathbf{c}) &= \frac{a_0}{1-\sum_{i=1}^q b_i}+\sum_{i=1}^p a_i x_{t-i}^2 \qquad (32)\\ &+\sum_{i=1}^p a_i\sum_{k=1}^\infty\sum_{j_i=1}^q\cdots\sum_{j_k=1}^q b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k}\ I_{t-i-j_1-\cdots j_k\geq 1}\end{aligned}$$

for $t\geq p+1$ and $\mathbf{c}=(\mathbf{a},\mathbf{b})$.

Then, from (31) and (32) we obtain

$$0\leq\widetilde{h}_t(\mathbf{c})-h_t(\mathbf{c})\leq\sum_{i=1}^p a_i\sum_{k=k_t}^\infty\sum_{j_i=1}^q\cdots\sum_{j_k=1}^q b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k},$$

where k_t is the integer part of $(t-p-1)/q$.

Define $\widetilde{b}=\max_C\{\max(b_1,...,b_q),(\mathbf{a},\mathbf{b})\in C\}$ and then

$$\sup_{\mathbf{c}\in C}\left|\widetilde{h}_t(\mathbf{c})-h_t(\mathbf{c})\right|\le \widetilde{b}^{k_t-1}\sup_{\mathbf{c}\in C}\sum_{i=1}^{p}a_i\sum_{k=1}^{\infty}\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q}b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k}.$$

Then (i) follows taking $\vartheta=\widetilde{b}^{1/q}$ and

$$W=\widetilde{b}^{-p/q-2}\sup_{(\mathbf{a},\mathbf{b})\in C}\sum_{i=1}^{p}a_i\sum_{k=1}^{\infty}\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q}b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k}.$$

Observe that $\sup_{(\mathbf{a},\mathbf{b})\in C}\sum_{i=1}^{q}b_i<1$ implies that $W<\infty$ and since $0<\widetilde{b}<1$ we also have $0<\vartheta<1$.

(ii) is proved in Hall and Yao (2003).

(iii) Hall and Yao (2003) derive the following formulas

$$\frac{\partial\widetilde{h}_t(\mathbf{c})}{\partial a_0}=\frac{1}{1-\sum_{j=1}^{q}b_j}, \tag{33}$$

$$\frac{\partial\widetilde{h}_t(\mathbf{c})}{\partial a_i}=x^2_{t-i}+\sum_{k=1}^{\infty}\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q}b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k},\ 1\le i\le p, \tag{34}$$

$$\begin{aligned}\frac{\partial\widetilde{h}_t(\mathbf{c})}{\partial b_j} &= \frac{a_0}{(1-\sum_{i=1}^{q}b_i)^2}\\ &\quad+\sum_{i=1}^{p}a_i\sum_{k=0}^{\infty}(k+1)\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q}b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k},\ 1\le j\le q\end{aligned} \tag{35}$$

In a similar way, from (32) we can derive

$$\frac{\partial h_t(\mathbf{c})}{\partial a_0}=\frac{1}{1-\sum_{j=1}^{q}b_j}, \tag{36}$$

$$\frac{\partial h_t(\mathbf{c})}{\partial a_i}=x^2_{t-i}+\sum_{k=1}^{\infty}\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q}b_{j_1}\cdots b_{j_k}x^2_{t-i-j_1-\cdots-j_k}I_{t-i-j_1-\cdots j_k\ge 1} \tag{37}$$

and

$$\frac{\partial h_t(\mathbf{c})}{\partial b_j} = \frac{a_0}{(1-\sum_{i=1}^{q} b_i)^2} + \tag{38}$$
$$+\sum_{i=1}^{p} a_i \sum_{k=0}^{\infty}(k+1)\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k} I_{t-i-j_1-\cdots j_k\geq 1}$$

for $t \geq p+1$ and $\mathbf{c}=(\mathbf{a},\mathbf{b})$. Then we get

$$\frac{\partial \widetilde{h}_t(\mathbf{c})}{\partial a_0} - \frac{\partial h_t(\mathbf{c})}{\partial a_0} = 0, \tag{39}$$

$$0 \leq \frac{\partial \widetilde{h}_t(\mathbf{c})}{\partial a_i} - \frac{\partial h_t(\mathbf{c})}{\partial a_i} \leq \sum_{k=k_t}^{\infty}\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k} \tag{40}$$

and

$$0 \leq \frac{\partial \widetilde{h}_t(\mathbf{c})}{\partial b_j} - \frac{\partial h_t(\mathbf{c})}{\partial b_j} \leq \sum_{i=1}^{p} a_i \sum_{k=k_t}^{\infty}(k+1)\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k}, \tag{41}$$

where k_t is the integer part of $(t-p-1)/q$. Consider a neighborhood U of $\boldsymbol{\gamma}$ such $U \subset C_{\delta_0/2}$, then, using a similar argument than the one used the proof of (i), we can prove that there exists ϑ_1, $0<\vartheta_1<1$, and a random variable W_1^* such that

$$\sup_{\mathbf{c}\in U}\left|\frac{\partial \widetilde{h}_t(\mathbf{c})}{\partial a_i} - \frac{\partial h_t(\mathbf{c})}{\partial a_i}\right| \leq \vartheta_1^t W_1^*. \tag{42}$$

A similar bound can be obtained for the right hand side of (41) as follows

$$\sum_{i=1}^{p} a_i \sum_{k=k_t}^{\infty}(k+1)\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k}$$
$$\leq \widetilde{b}^{k_t-1}\sum_{i=1}^{p} a_i \sum_{k=1}^{\infty}(k_t+k)\sum_{j_i=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k},$$

where $\widetilde{b} = \max_C \{\max(b_1,...,b_q),(\mathbf{a},\mathbf{b}) \in C\}$.

There exists $0 < b^* < 1$ and t_0 such that for $t \geq t_0$ we have

$$\widetilde{b}^{k_t - 1} k_t \leq \widetilde{b}^{t/q} t \widetilde{b}^{-p/q-2} \leq (b^*)^t \, \widetilde{b}^{-p/q-2}.$$

and from Hall and Yao (2003) we know that the random variables

$$\sup_{\mathbf{c}\in U} \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x^2_{t-i-j-j_1-\cdots-j_k}$$

and

$$\sup_{\mathbf{c}\in U} \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} k \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x^2_{t-i-j-j_1-\cdots-j_k}$$

are finite. Then taking $\vartheta_2 = b^*$ and

$$W_2^* = \widetilde{b}^{-p/q-2} \sup_{(\mathbf{a},\mathbf{b})\in C} \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} (1+k) \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x^2_{t-i-j_1-\cdots-j_k}$$

we obtain

$$\sup_{\mathbf{c}\in U} \left| \frac{\partial \widetilde{h}_t(\mathbf{c})}{\partial b_j} - \frac{\partial h_t(\mathbf{c})}{\partial b_j} \right| \leq \vartheta_2^t W_2^* \tag{43}$$

for $t \geq t_0$ where W_2^* is a finite random variable.

From (39),(42) and (43) we get that there exists a constant $0 < \vartheta_3 < 1$ and a finite random variable W_3^* such that

$$\sup_{\mathbf{c}\in U} \left| \nabla \widetilde{h}_t(\mathbf{c}) - \nabla h_t(\mathbf{c}) \right| \leq \vartheta_3^t W_3^* \tag{44}$$

for $t \geq t_0$

Since $\mathbf{c} \in U \subset C_{\delta_0/2}$, from (21) we obtain

$$\widetilde{h}_t(\mathbf{c}) \geq \delta_0/2, \;\; h_t(\mathbf{c}) \geq \delta_0/2. \tag{45}$$

We can write

$$\begin{aligned} & \nabla \log \widetilde{h}_t(\mathbf{c}) - \nabla \log h_t(\mathbf{c}) \\ = \;& \frac{1}{h_t(\mathbf{c})\widetilde{h}_t(\mathbf{c})} \nabla \widetilde{h}_t(\mathbf{c}) \left(h_t(\mathbf{c}) - \widetilde{h}_t(\mathbf{c}) \right) + \frac{1}{h_t(\mathbf{c})} (\nabla \widetilde{h}_t(\mathbf{c}) - \nabla h_t(\mathbf{c})), \end{aligned}$$

and so from (i), (ii), (44) and (45) we prove (iii).

(iv) We can write

$$\begin{aligned}
& \nabla \log(\widetilde{h}_t(\mathbf{c}))\nabla \log(\widetilde{h}_t(\mathbf{c}))' - \nabla \log(h_t(\mathbf{c}))\nabla \log(h_t(\mathbf{c}))' \\
= \; & \nabla \log(\widetilde{h}_t(\mathbf{c}))(\nabla \log(\widetilde{h}_t(\mathbf{c})) - \nabla \log(h_t(\mathbf{c})))' + \left(\nabla \log(\widetilde{h}_t(\mathbf{c})) - \nabla \log(h_t(\mathbf{c}))\right) \nabla \log(h_t(\mathbf{c}))'
\end{aligned}$$

Then using (ii) and (iii) we get (iv).

(v) This is shown by Peng and Yao (2003) while proving Theorem 1.

(vi) The proof is similar to that of (iv) using the expression for $\nabla^2 \widetilde{h}_t(\mathbf{c})$ given in Peng and Yao (2003).

Lemma 3. *Let C be as in (21) and*

$$\widetilde{M_T}(\mathbf{c}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log \widetilde{h}_t(\mathbf{c})). \tag{46}$$

Then under the assumptions of Theorem 2, have

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in C} \left| \widetilde{M_T}(\mathbf{c}) - M(\mathbf{c}) \right| = 0 \; a.s..$$

Proof.Proof.

We start proving

$$E\left(\sup_{\mathbf{c}\in\mathbf{C}} (|\rho(y_t - \log \widetilde{h}_t(\mathbf{c}))|) \right) < \infty. \tag{47}$$

Since ρ has a bounded derivative, it is enough to show that

$$E\left(\sup_{\mathbf{c}\in\mathbf{C}} (|y_t - \log \widetilde{h}_t(\mathbf{c})|) \right) < \infty. \tag{48}$$

We also have

$$y_t - \log \widetilde{h}_t(\mathbf{c}) = w_t + \log \widetilde{h}_t(\boldsymbol{\gamma}) - \log(\widetilde{h}_t(\mathbf{c})).$$

Since by Lemma 1 (a) $E(|w_t|) < \infty,$ to prove (47) it is enough to show that

$$E\left(\sup_{\mathbf{c}\in\mathbf{C}} (|\log \widetilde{h}_t(\mathbf{c})|) \right) < \infty. \tag{49}$$

From (21) and (31) we have that

$$E(\sup_{\mathbf{c}\in C} \widetilde{h}_t(\mathbf{c})) < \infty, \tag{50}$$

and from (21) we obtain

$$\delta_0 \leq \inf_{\mathbf{c}\in C} \widetilde{h}_t(\mathbf{c}). \tag{51}$$

Then (49) follows from (50) and (51) and the Lemma follows from Lemma 3 of Muler and Yohai (2002).

Lemma 4. *Under the assumptions of Theorem 2, we have*

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in C} \left| M_T(\mathbf{c}) - \widetilde{M}_T(\mathbf{c}) \right| = 0$$

a.s.

Proof

We have

$$M_T(\mathbf{c}) - \widetilde{M}_T(\mathbf{c}) = \frac{1}{(T-p)} \sum_{t=p+1}^{T} \left(\rho\left(y_t - \log\left(\widetilde{h}_t(\mathbf{c}) \right) \right) - \rho\left(y_t - \log\left(h_t(\mathbf{c}) \right) \right) \right).$$

Let $K = \sup |\rho'| < \infty$. Then, from Lemma 2-(i) and (21) we have that there exists $0 < \vartheta < 1$, and a finite positive random variable W such that

$$\left| \rho\left(y_t - \log\left(\widetilde{h}_t(\mathbf{c}) \right) \right) - \rho\left(y_t - \log\left(h_t(\mathbf{c}) \right) \right) \right| \leq \frac{K}{\delta_0} \left| \widetilde{h}_t(\mathbf{c}) - h_t(\mathbf{c}) \right| \leq \frac{K}{\delta_0} \vartheta^t W,$$

and this proves the Lemma.

Proof of Theorem 2.

From Lemmas 3 and 4 we get

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in C} |M_T(\mathbf{c}) - M(\mathbf{c})| = 0$$

a.s. Then, putting

$$A = \left\{ \sup_{\mathbf{c}\in C} |M_T(\mathbf{c}) - M(\mathbf{c})| \to 0 \right\},$$

we have $P(A) = 1$. Therefore it is enough to prove

$$A \subset \{ \widehat{\gamma}_T \to \gamma \}. \tag{52}$$

Assume that (52) is not true. Then we can find in A a subsequence $\widehat{\boldsymbol{\gamma}}_{T_i}$ such that

$$\widehat{\boldsymbol{\gamma}}_{T_i} \to \overline{\boldsymbol{\gamma}} \tag{53}$$

with $\overline{\boldsymbol{\gamma}} \neq \boldsymbol{\gamma}$. Since $M(\mathbf{c})$ has a unique minimum at $\boldsymbol{\gamma}$, and $M(\mathbf{c})$ is continuous, there exists a neighborhood $U(\overline{\boldsymbol{\gamma}})$ and $\varepsilon > 0$ such that for all $\mathbf{c} \in U(\overline{\boldsymbol{\gamma}})$ we obtain

$$M(\mathbf{c}) > M(\boldsymbol{\gamma}) + \varepsilon. \tag{54}$$

From (53) there exists i_0 large enough such that for all $i \geq i_0$ we obtain

$$\widehat{\boldsymbol{\gamma}}_{T_i} \in U(\overline{\boldsymbol{\gamma}}), \sup_{\mathbf{c}\in\mathbf{C}} |M_{T_i}(\mathbf{c}) - M(\mathbf{c})| < \frac{\varepsilon}{2}. \tag{55}$$

Therefore from (54) and (55) for all $i \geq i_0$ we obtain

$$M_{T_i}(\widehat{\boldsymbol{\gamma}}_{T_i}) = M_{T_i}(\widehat{\boldsymbol{\gamma}}_{T_i}) - M(\widehat{\boldsymbol{\gamma}}_{T_i}) + M(\widehat{\boldsymbol{\gamma}}_{T_i}) > M(\boldsymbol{\gamma}) + \frac{\varepsilon}{2}. \tag{56}$$

Using the definition of $\widehat{\boldsymbol{\gamma}}_T$ and (55), we have

$$M_{T_i}(\widehat{\boldsymbol{\gamma}}_{T_i}) \leq M_{T_i}(\boldsymbol{\gamma}) < M(\boldsymbol{\gamma}) + \frac{\varepsilon}{2},$$

for all $i \geq i_0$. This contradicts (56) and therefore the Theorem is proved.

We need the following four Lemmas to prove Theorem 3.

Lemma 5. *Suppose that all the assumptions of Theorem 2 hold. Moreover, assume that ρ has a continuous and bounded derivative ψ such that $E(\psi^2(w_t)) > 0$. Then,*

$$\frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \nabla\rho\left(y_t - \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right)\right) \to_D N(\mathbf{0}, E(\psi^2(w_t))D_0),$$

where

$$D_0 = E\left(\nabla \log(\widetilde{h}_t(\boldsymbol{\gamma}))\nabla \log(\widetilde{h}_t(\boldsymbol{\gamma}))'\right).$$

Proof.

From Lemma 2-(ii) D_0 is finite, and from (4) it can be shown that D_0 is positive definite (see for instance Horvath and Kokoszka (2003)).

On the other hand, since $E(\rho(w_t - u))$ is minimized at $u = 0$, we have

$$E(\psi(w_t)) = 0.$$

This implies that $\mathbf{b}'\psi(w_t\,)\nabla\log(\widetilde{h}_t(\boldsymbol{\gamma}))$ is a stationary martingale difference sequence for any vector $\mathbf{b} \neq \mathbf{0}$ in R^{p+q+1}. Then applying the Central Limit Theorem for Martingales (see for instance Theorem 24.4, Davidson(1994)) we obtain

$$\frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \mathbf{b}'\psi(w_t)\nabla\log(\widetilde{h}_t(\boldsymbol{\gamma})) \ \rightarrow_D N(0, E(\psi^2\,(w_t))\mathbf{b}'\,D_0\mathbf{b}).$$

Finally, using a standard Cramer-Wold device we get the desired result.

Lemma 6. *Suppose that all the assumptions of Lemma 5 hold. Moreover, assume that ρ has a two continuous and bounded derivatives and that $E(\psi'(w_t)) > 0$. Define $A(\mathbf{c}) = E(\nabla^2\rho(y_t - \log(\widetilde{h}_t(\mathbf{c}))))$, then there exists a neighborhood U of $\boldsymbol{\gamma}$ such that*

(i)

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in\mathbf{U}} \left\| \frac{1}{T-p} \sum_{t=p+1}^{T} \nabla^2\rho(y_t - \log(\widetilde{h}_t(\mathbf{c}))) - A(\mathbf{c}) \right\| = 0 \text{ a.s..}$$

(ii) *$A(\boldsymbol{\gamma})$ is a positive definite matrix given by $A(\boldsymbol{\gamma}) = E(\psi'(w_t))D_0$.*

Proof.

Differentiating $\nabla\rho\left(y_t - \log\left(\widetilde{h}_t(\mathbf{c})\right)\right)$ we get

$$\begin{aligned} & \nabla^2\rho\left(y_t - \log\left(\widetilde{h}_t(\mathbf{c})\right)\right) \\ = \;& (\psi'(y_t - \log\widetilde{h}_t(\mathbf{c})) + \psi(y_t - \log(\widetilde{h}_t(\mathbf{c})))\nabla\log(\widetilde{h}_t(\mathbf{c}))\nabla\log\left(\widetilde{h}_t(\mathbf{c})\right)' \\ & -\psi(y_t - \log(\widetilde{h}_t(\mathbf{c})))\frac{\nabla^2\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})}. \end{aligned} \tag{57}$$

From (21) and Lemma 2-(v) there exists a neighborhood U of $\boldsymbol{\gamma}$ such that

$$E\left(\sup_{\mathbf{c}\in U}\left\|\frac{\nabla^2\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})}\right\|\right) < \infty. \tag{58}$$

Then by Lemma 2-(ii), (57), (58) and the fact that ψ and ψ' are continuous and bounded we get

$$E \sup_{\mathbf{c} \in U} \left\| \nabla^2 \rho \left(y_t - \log \left(\widetilde{h}_t(\mathbf{c}) \right) \right) \right\| < \infty.$$

Therefore, part (i) of the Lemma follows from Lemma 3 of Muler and Yohai (2002).

Since $E(\psi(w_t)) = 0$ and $y_t - \log(\widetilde{h}_t(\boldsymbol{\gamma})) = w_t$ we get

$$E(\psi(y_t - \log(\widetilde{h}_t(\boldsymbol{\gamma}))\nabla \log(\widetilde{h}_t(\boldsymbol{\gamma}))\nabla \log(\widetilde{h}_t(\boldsymbol{\gamma}))) = 0$$

and

$$E\left(\psi(y_t - \log(\widetilde{h}_t(\boldsymbol{\gamma})))\frac{\nabla^2 \widetilde{h}_t(\boldsymbol{\gamma})}{\widetilde{h}_t(\boldsymbol{\gamma})}\right) = 0.$$

Then

$$A(\boldsymbol{\gamma}) = E(\psi'(w_t))D_0.$$

Since D_0 is positive definite and $E(\psi'(w_t)) > 0$ part (ii) follows.

Lemmas 7 and 8 are necessary to show that the asymptotic distribution of the M-estimates can be derived using the $\widetilde{h}_t(\boldsymbol{\gamma})$'s instead of the $h_t(\boldsymbol{\gamma})$'s.

Lemma 7. *Suppose that all the assumptions of Lemma 6 hold. Then*

$$\lim_{T\to\infty} \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \left\| \nabla \rho \left(y_t - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) - \nabla \rho \left(y_t - \log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right) \right) \right\| = 0.$$

a.s..

Proof.

We can write

$$\begin{aligned} & \nabla \rho \left(y_t - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) - \nabla \rho \left(y_t - \log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right) \right) \\ = \;& \psi \left(y_t - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) \nabla \left(\log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right) - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) \\ & + \left(\psi \left(y_t - \log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right) \right) - \psi \left(y_t - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) \right) \nabla \log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right). \end{aligned} \quad (59)$$

From Lemma 2-(iii) there exists a finite and postive random variable W_1 and a constant $0 < \vartheta < 1$ such that

$$\lim_{T\to\infty} \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \left\| \nabla \left(\log \left(\widetilde{h}_t(\boldsymbol{\gamma}) \right) - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) \right\| \leq \lim_{T\to\infty} \frac{W_1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \vartheta^t = 0$$

a.s.. Then, since ψ is bounded we have

$$\lim_{T\to\infty} \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \Big\| \psi\left(y_t - \log(h_t(\boldsymbol{\gamma}))\right) \Big(\nabla \log(\widetilde{h}_t(\boldsymbol{\gamma})) - \nabla \log(h_t(\boldsymbol{\gamma}))\Big)\Big\| = 0 \tag{60}$$

a.s..

The function ψ' is bounded and both $\widetilde{h}_t(\boldsymbol{\gamma})$ and $h_t(\boldsymbol{\gamma})$ has positive lower bounds, so using the Mean Value Theorem we have that there exists a constant $k_1 > 0$ such that

$$\frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \left\| \left(\psi\left(y_t - \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right)\right) - \psi\left(y_t - \log\left(h_t(\boldsymbol{\gamma})\right)\right)\right) \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\|$$

$$\leq \frac{k_1}{\sqrt{T-p}} \left(\sum_{t=p+1}^{T} \left| \widetilde{h}_t(\boldsymbol{\gamma}) - h_t(\boldsymbol{\gamma}) \right| \left\| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\| \right). \tag{61}$$

By Lemma 2-(i) there exists a finite positive random variable W and $0 < \vartheta < 1$ such that

$$\sum_{t=p+1}^{T} \left| \widetilde{h}_t(\boldsymbol{\gamma}) - h_t(\boldsymbol{\gamma}) \right| \left\| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\| \leq W \sum_{t=p+1}^{T} \vartheta^t \left\| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\|. \tag{62}$$

Define for all $T \geq p+1$,

$$S_T = \sum_{t=p+1}^{T} \vartheta^t \left\| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\|,$$

Since from Lemma 2-(ii), $E\left(\left| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right| \right) < \infty$, we get $\lim_{T\to\infty} E(S_T) < \infty$ and then

$$S = \sum_{t=p+1}^{\infty} \vartheta^t \left\| \nabla \log\left(\widetilde{h}_t(\boldsymbol{\gamma})\right) \right\| \tag{63}$$

is a finite random variable with $E(S) < \infty$. Then from (61), (62) and (63) we obtain

$$\lim_{T\to\infty} \frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \left\| \left(\psi \left(y_t - \log(\widetilde{h}_t(\boldsymbol{\gamma})) \right) \right.\right.$$

$$\left.\left. -\psi\left(y_t - \log(h_t(\boldsymbol{\gamma}))\right)\right) \nabla \log(\widetilde{h}_t(\boldsymbol{\gamma})) \right\| = \lim_{T\to\infty} \frac{k_1 W S}{\sqrt{T-p}} = 0 \tag{64}$$

a.s..

Then, the lemma follows from (59), (60) and (64).

Lemma 8. *Suppose that all the assumptions of Theorem 3 hold. Then, there exists a neighborhood U of $\boldsymbol{\gamma}$ such that*

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in U} \frac{1}{T-p} \left\| \sum_{t=p+1}^{T} \nabla^2 \rho \left(y_t - \log\left(\widetilde{h}_t(\mathbf{c})\right)\right) - \nabla^2 \rho \left(y_t - \log\left(h_t(\mathbf{c})\right)\right) \right\| = 0$$

a.s..

Proof.

Let $\zeta(x) = \left(\psi'(x) + \psi(x)\right)$. We can write from (57),

$$\nabla^2 \rho \left(y_t - \log\left(\widetilde{h}_t(\mathbf{c})\right)\right) - \nabla^2 \rho \left(y_t - \log\left(h_t(\mathbf{c})\right)\right) = H_t(\mathbf{c}) - G_t(\mathbf{c}), \tag{65}$$

where

$$\begin{aligned} H_t(\mathbf{c}) &= \zeta(y_t - \log(\widetilde{h}_t(\mathbf{c}))) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right)' \\ &\quad -\zeta(y_t - \log(h_t(\mathbf{c}))) \nabla \log\left(h_t(\mathbf{c})\right) \nabla \log\left(h_t(\mathbf{c})\right)' \end{aligned}$$

and

$$\begin{aligned} G_t(\mathbf{c}) &= \psi\left(y_t - \log\left(\widetilde{h}_t(\mathbf{c})\right)\right) \frac{\nabla^2 \widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})} \\ &\quad -\psi\left(y_t - \log\left(h_t(\mathbf{c})\right)\right) \frac{\nabla^2 h_t(\mathbf{c})}{h_t(\mathbf{c})}. \end{aligned} \tag{66}$$

Let

$$Q_t(\mathbf{c}) = \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right)' - \nabla \log\left(h_t(\mathbf{c})\right) \nabla \log\left(h_t(\mathbf{c})\right)'.$$

Then, we have

$$H_t(\mathbf{c}) = \zeta(y_t - \log(h_t(\mathbf{c})))Q_t(\mathbf{c}) \tag{67}$$
$$+ \left(\zeta(y_t - \log(\widetilde{h}_t(\mathbf{c}))) - \zeta(y_t - \log(h_t(\mathbf{c})))\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right)'.$$

From Lemma 2-(iv) there exists a neighborhood U_1 of $\boldsymbol{\gamma}$ with $U_1 \subset C_{\delta_0/2}, 0 < \vartheta < 1$ and a finite random variable W_2 such that

$$\sup_{\mathbf{c}\in U_1} \|Q_t(\mathbf{c})\| < \vartheta^t W_2$$

for $t \geq p+1$. Then

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in U_1} \frac{1}{T-p} \sum_{t=p+1}^{T} \|Q_t(\mathbf{c})\| = 0 \tag{68}$$

a.s.. Then, since $\zeta(y_t - \log(h_t(\mathbf{c})))$ is bounded we obtain

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in U_1} \frac{1}{T-p} \sum_{t=p+1}^{T} \|\zeta(y_t - \log(h_t(\mathbf{c})))Q_t(\mathbf{c})\| = 0 \tag{69}$$

a.s.. From Lemma 2-(ii) we have that there exists a neighborhood $U_2 \subset U_1$ of $\boldsymbol{\gamma}$ such that,

$$E\left(\sup_{\mathbf{c}\in U_2} \left\|\nabla \log\left(\widetilde{h}_t(\mathbf{c})\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right)'\right\|\right) < \infty. \tag{70}$$

Since ζ' is bounded and $U_2 \subset C_{\delta_0/\,2}$, applying Lemma 2-(i) and using similar arguments that in the proof of Lemma 7, we obtain

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in U_2} \frac{1}{T-p} \sum_{t=p+1}^{T} \left(\zeta(y_t - \log(\widetilde{h}_t(\mathbf{c}))) - \zeta(y_t - \log(h_t(\mathbf{c})))\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right) \nabla \log\left(\widetilde{h}_t(\mathbf{c})\right)' = 0 \tag{71}$$

a.s.. Then, by (67), (69) and (71) we have

$$\lim_{T\to\infty} \sup_{\mathbf{c}\in U_2} \frac{1}{T-p} \sum_{t=p+1}^{T} H_t(\mathbf{c}) = 0 \tag{72}$$

a.s..

From Lemma 2-(vi) and Lemma 2-(v) there exists a neighborhood $U \subset U_2$ of $\boldsymbol{\gamma}$, a constant $0 < \vartheta < 1$ and a positive variable W_3 such that for all $t \geq p+1$

$$\sup_{\mathbf{c} \in U} \left\| \frac{\nabla^2 \widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})} - \frac{\nabla^2 h_t(\mathbf{c})}{h_t(\mathbf{c})} \right\| \leq \vartheta^t W_3,$$

and

$$\sup_{\mathbf{c} \in U} E \left\| \frac{\nabla^2 \widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})} \right\|^2 < \infty.$$

Then, since ψ' is bounded and $U \subset C_{\delta_0/2}$ using similar arguments than in the proof of Lemma 7, we can show

$$\lim_{T \to \infty} \sup_{\mathbf{c} \in U} \frac{1}{T-p} \sum_{t=p+1}^{T} G_t(\mathbf{c}) = 0 \tag{73}$$

a.s.. Finally, the Lemma follows from (72) and (73).

Proof of Theorem 3

From Lemmas 5 and 7 we have

$$\frac{1}{\sqrt{T-p}} \sum_{t=p+1}^{T} \nabla \rho \left(y_t - \log \left(h_t(\boldsymbol{\gamma}) \right) \right) \ \to_D N(\mathbf{0}, E(\psi^2 \left(w_t \right)) D_0), \tag{74}$$

and from Lemmas 6-(i) and 8 we get that there exists a neighborhood U of $\boldsymbol{\gamma}$ such that

$$\lim_{T \to \infty} \sup_{\mathbf{c} \in U} \left\| \frac{1}{T-p} \sum_{t=p+1}^{T} \nabla^2 \rho \left(y_t - \log \left(h_t(\mathbf{c}) \right) \right) - A(\mathbf{c}) \right\| = 0 \text{ a.s..} \tag{75}$$

From (74) and (75) and Theorem 2 we get that

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \nabla^2 \rho \left(y_t - \log \left(h_t(\mathbf{c}) \right) \right)$$

is continuous in $\mathbf{c}$, and that $A_0 = A(\boldsymbol{\gamma})$ is nonsingular (Lemma 6-(ii)). Then, Theorem 3 follows from Theorem 4.1.3. of Amemiya (1985).

The following Lemmas 9, 10 and 11 are going to be used to prove Theorem 4.

Lemma 9. *Let* $\mathbf{Y}_t$ *be an ergodic process in* R^m *and* $g : R^m \times R \rightarrow R$, *a continuous function satisfying:*

(i) *There exists* $g_0 : R^m \rightarrow R$ *such that* $|g(\mathbf{Y}_t, u)| \leq g_0(\mathbf{Y}_t)$, *and* $g_0(\mathbf{Y}_t)$ *integrable.*

(ii) $\lim_{u\rightarrow\infty} g(\mathbf{Y}_t, u) = g^+(\mathbf{Y}_t)$ *and* $\lim_{u\rightarrow-\infty} g(\mathbf{Y}_t, u) = g^-(\mathbf{Y}_t)$. *Then*

$$\lim_{T\rightarrow\infty} \sup_{u\in\ R} \left| \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \right| = 0 \text{ a.s..}$$

Proof.

From Muler and Yohai (2002) we have

$$\lim_{T\rightarrow\infty} \sup_{u\in\ K} \left| \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \right| = 0 \text{ a.s.} \tag{76}$$

for any compact set $K \subset R$. Then to prove the Lemma, it is enough to show that given any ε, there exists $\overline{u}$ such that

$$\lim_{T\rightarrow\infty} \sup_{u\geq\overline{u}} \left| \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \right| \leq \varepsilon \text{ a.s.} \tag{77}$$

and

$$\lim_{T\rightarrow\infty} \sup_{u\leq\underline{u}} \left| \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \right| \leq \varepsilon \text{ a.s..}$$

Since both proofs are similar we only show (77). To this purpose, it is enough to prove that

$$\lim_{T\rightarrow\infty} \sup_{u\geq\overline{u}} \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \leq \varepsilon \text{ a.s.} \tag{78}$$

and

$$\lim_{T\rightarrow\infty} \inf_{u\geq\overline{u}} \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \geq -\varepsilon \text{ a.s..} \tag{79}$$

Since the proofs of (78) and (79)are similar, we only show (78).

Let

$$B_t(v) = \sup_{u>v}(g(\mathbf{Y}_t, u) - E(g(\mathbf{Y}_t, u)).$$

Clearly, by the Dominated Convergence Theorem we have $\lim_{v\to\infty} B_t(v) = g^+(\mathbf{Y}_t) - E(g^+(\mathbf{Y}_t))$ and $\lim_{v\to\infty} E(B_t(v)) = 0$. Therefore there exists $\overline{u}$ such that $E(B_t(\overline{u})) < \varepsilon$. Then using the Law of Large Numbers we get

$$\lim_{T\to\infty} \sup_{u\geq\overline{u}} \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E\left(g(\mathbf{Y}_t, u)\right) \leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} B_t(\overline{u}) \leq \varepsilon \text{ a.s.}$$

and this proves (78).

Lemma 10. *Suppose that all the assumptions of Theorem 2 hold. Then, we have that*

$$\sup_{\mathbf{c}\in C} h^*_{t,k}(\mathbf{c}) \leq R_t,$$

where $h^*_{t,k}(\mathbf{c})$ *is defined in (15) and* R_t *is a positive-valued ergodic process.*

Proof.

Define

$$R_t = \sup_{\mathbf{c}\in\mathbf{C}} \widetilde{h}_t(\mathbf{c}). \tag{80}$$

Then, from (21) and (31) we have that R_t is a positive-valued ergodic processes and from (8) we get

$$\sup_{\mathbf{c}\in C} h_t(\mathbf{c}) \leq R_t. \tag{81}$$

We prove by induction on t that

$$h^*_{t,k}(\mathbf{c}) \leq h_t(\mathbf{c}). \tag{82}$$

From the (15) it follows immediately that

$$h^*_{t,k}(\mathbf{c}) = h_t(\mathbf{c}) \tag{83}$$

for all $t \leq 0$. Assume now

$$h^*_{j,k}(\mathbf{c}) \leq h_j(\mathbf{c}),\ j \leq t.$$

Then from (15) we have

$$h^*_{t+1,k}(\mathbf{c}) \leq a_0 + \sum_{i=1}^{p} a_i x^2_{t+1-i} + \sum_{i=1}^{q} b_i h^*_{t+1-i,k}(\mathbf{c}) \leq \ h_{t+1}(\mathbf{c}), \tag{84}$$

and thereby (82) follows. Then, the Lemma follows from (81).

Lemma 11. *Suppose that all the assumptions of Theorem 4 hold. Let* $m_0 = E(\rho(w_t)) = J(0)$. *Then, there exists* $\delta > 0$ *such that*

$$\liminf_{T\to\infty} \inf_{\mathbf{c}\in C} M^*_{Tk}(\mathbf{c}) > m_0 + \delta \text{ a.s.},$$

where M^*_{Tk} *is given in (18).*

Proof.

Since $\boldsymbol{\gamma} \in C$, there exists i_0, $1 \leq i_0 \leq p$ such that $\alpha_{i_0} > 0$. Then

$$\widetilde{h}_t(\boldsymbol{\gamma}) \geq \alpha_{i_0} x^2_{t-i_0} \ = \alpha_{i_0} z^2_{t-i_0} \widetilde{h}_{t-i_0}(\boldsymbol{\gamma}). \tag{85}$$

Consider $s = \max(p,q)$. If $p < s$ define $a_{p+1} = \cdots = a_s = 0$ and $\alpha_{p+1} = \cdots = \alpha_s = 0$. If $q < s$ define $b_{q+1} = \cdots = b_s = 0$ and $\beta_{q+1} = \cdots = \beta_s = 0$. Then, we have for all $t \geq 1$

$$h^*_{t,k}(\mathbf{c}) \leq a_0 + \sum_{i=1}^{s} (a_i k + b_i) h^*_{t-i,k}(\mathbf{c}) \leq (2+k) \sum_{i=1}^{s} h^*_{t-i,k}(\mathbf{c}) \tag{86}$$

and

$$\sum_{i=1}^{s} h^*_{t-i,k}(\mathbf{c}) \leq (2+k) \sum_{i=1}^{s} h^*_{t-i-1,k}(\mathbf{c}) + \sum_{i=2}^{s} h^*_{t-i,k}(\mathbf{c}) \leq 2(2+k) \sum_{i=1}^{s} h^*_{t-i-1,k}(\mathbf{c}). \tag{87}$$

By Lemma 10, there exists a positive-valued ergodic process R_t such that $\sup_{\mathbf{c}\in C} h^*_{t,k}(\mathbf{c}) \leq R_t$. Then, from (86) and (87) we have that

$$h^*_{t,k}(\mathbf{c}) \leq 2^{i_0} (2+k)^{i_0+1} \sum_{i=j+1}^{s+j} R_{t-i}. \tag{88}$$

Let us define the ergodic processes

$$N_t = \frac{\widetilde{h}_{t-i_0}(\boldsymbol{\gamma})}{\sum_{i=i_0+1}^{s+i_0} R_{t-i}}, t \geq 1.$$

Then there exists $\eta > 0$ and $\nu > 0$ such that

$$P(N_t > \eta\ ) \geq \nu. \tag{89}$$

Using that $\lim_{u\to\infty} \rho(u) = \sup_u \rho(u) > m_{0,}$ it is easy to show that if ε is small enough, there exists $k_1 > 0$ such that

$$\inf_{u\leq -k_1} E(\rho(w_t - u)) \geq m_0 + \varepsilon. \tag{90}$$

Let us define K as

$$K = \frac{2^{i_0}(2+k)^{i_0+1}e^{k_1}}{\eta\alpha_{i_0}} \tag{91}$$

and define

$$A_t = \left\{N_t > \eta,\ z_{t-i_0}^2 > K\right\}. \tag{92}$$

Since $z_{t-i_0}^2$ is independent of N_t, using (89) and the fact that z_t is unbounded, we have

$$a = P(A_t) = \nu P(z_{t-i_0}^2 > K) > 0. \tag{93}$$

From (88), (85) and the choice of K in the definition of A_t we have

$$\sup_{\mathbf{c}\in C}(\log h_{t,k}^*(\mathbf{c}) - \log \widetilde{h}_t(\boldsymbol{\gamma})) \leq -k_1. \tag{94}$$

We can write

$$M_{Tk}^*(\mathbf{c}) = \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t + \log(\widetilde{h}_t(\boldsymbol{\gamma})) - \log(h_{t,k}^*(\mathbf{c}))).$$

Let us consider first the case when ρ is a bounded function. From (94) we obtain

$$\begin{aligned}\inf_{\mathbf{c}\in C} M_{Tk}^*(\mathbf{c}) \quad &\geq \quad \inf_u \frac{1}{T-p}\sum_{t=p+1}^{T}\rho(w_t - u)(1 - I_{A_t}) \\ &\quad + \inf_{u\leq -k_1}\frac{1}{T-p}\sum_{t=p+1}^{T}\rho(w_t - u)I_{A_t}. \end{aligned} \tag{95}$$

Since ρ is bounded, $\lim_{|u|\to\infty}\rho(u) = \sup_u \rho(u)$ and I_{A_t} is ergodic and independent of w_t, from Lemma 9 we get

$$\lim_{T\to\infty} \sup_u \left| \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t}) - E\left(\rho(w_t-u)\right)(1-a) \right| = 0 \text{ a.s.} \tag{96}$$

and

$$\lim_{T\to\infty} \sup_{u\le -k_1} \left| \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(w_t-u)I_{A_t} - E\left(\rho(w_t-u)\right)a \right| = 0 \text{ a.s..} \tag{97}$$

From (96) and using $E\left(\rho(w_t-u)\right) \ge m_0 > 0$ we get

$$\liminf_{T\to\infty} \inf_u \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t}) \ge m_0(1-a) \text{ a.s.},$$

and from (90) and (97) we obtain

$$\liminf_{T\to\infty} \inf_{u\le -k_1} \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(w_t-u)I_{A_t} \ge (m_0+\varepsilon)a \text{ a.s..}$$

Thus, from (95) we derive

$$\liminf_{T\to\infty} \inf_{\mathbf{c}\in C} M_{Tk}^*(\mathbf{c}) \ge m_0 + \varepsilon a \text{ a.s.},$$

and then taking $\delta = \varepsilon a$, the Lemma follows for the case that ρ is bounded.

Consider now the case that $\lim_{|x|\to\infty} \rho(x) = \infty$. We start proving that for any $k \ge 0$,

$$E(\sup_{|u|\le k} \rho(w_t-u)) < \infty. \tag{98}$$

Take any sequence u_i with $|u_i| \le k$, then

$$\begin{aligned} \limsup_{i\to+\infty} E(\rho(w_t-u_i)) &= \limsup_{i\to+\infty} E\left(\rho(w_t-u_i)-\rho(w_t)\right) + E(\rho(w_t)) \\ &\le k\sup_u \rho'(u) + E(\rho(w_t)), \end{aligned}$$

and (98) is proved.

We will prove that there exists k_1 large enough such that

$$\liminf_{T\to\infty} \sup_{u\le -k_1} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)I_{A_t} \ge 2m_0 a \text{ a.s.} \tag{99}$$

Given $D>0$, define $B_{tD}=\{|w_t|\le D\}$ and let D_0 be such $P(B_{tD_0})\ge 1/2$. Let d_1 be such that for all $|x|\ge d_1$ we have $\rho(x)\ge 4m_0$ and put $k_1=2D_0+d_1$. Then,

$$\begin{aligned}
& \liminf_{T\to\infty} \sup_{u\le -k_1} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)I_{A_t} \\
\ge\ & \liminf_{T\to\infty} \sup_{u\le -k_1} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)I_{A_t}I_{B_{tD_0}} \\
\ge\ & 4m_0 \liminf_{T\to\infty} \frac{1}{T-p}\sum_{t=p+1}^{T} I_{A_t}I_{B_{tD_0}} = 4m_0 E(I_{A_t}I_{B_{tD_0}}).
\end{aligned}$$

Since A_t and B_t are independent, $P(B_{tD_0})\ge 1/2$ we obtain (99).

We prove now

$$\lim \inf_{T\to\infty} \inf_{u} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t}) \ge m_0\,(1-a) \;\text{ a.s..} \tag{100}$$

To prove this it is enough to show that

$$\lim_{D\to\infty} \lim \inf_{T\to\infty} \inf_{u} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t})I_{B_{tD}} \ge m_0\,(1-a) \;\text{ a.s..} \tag{101}$$

Since $\lim_{|x|\to\infty}\rho(x)=\infty$, there exists a compact set $U_D\subset R$ such that

$$\inf_{u} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t})I_{B_{tD}} = \inf_{u\in U_D} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t})I_{B_{tD}},$$

and so using compacity arguments and the fact that I_{A_t} is independent of w_t and of B_{tD} we have that

$$\lim \inf_{T\to\infty} \inf_{u\in U_D} \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(w_t-u)(1-I_{A_t})I_{B_{tD}}$$

$$\begin{aligned} &\geq \inf_{u \in U_D} E\left(\rho(w_t - u) I_{B_{tD}}\right)(1-a) \\ &\geq \inf_{u} E\left(\rho(w_t - u) I_{B_{tD}}\right)(1-a). \end{aligned}$$

Then to prove (100) it is enough to show that

$$\lim_{D \to \infty} \inf_{u} E\left(\rho(w_t - u) I_{B_{tD}}\right) = m_0. \tag{102}$$

Suppose that (102) is not true. Then, there exists (u_n, D_n) such that $D_n \uparrow \infty$ and $\varepsilon > 0$ such that

$$E\left(\rho(w_t - u_n) I_{B_{tD_n}}\right) < m_0 - \varepsilon. \tag{103}$$

Without loss of generality, taking if necessary a subsequence, we have to consider only the following two cases

(i) $\lim_{n \to \infty} u_n = \overline{u}$.

(ii) $\lim_{n \to \infty} |u_n| = \infty$.

We have that

$$\lim_{n \to \infty} \rho(w_t - u_n) I_{B_{tD_n}} = \rho(w_t - \overline{u})$$

Since

$$\sup_{|u| \leq 2\overline{|u|}} \rho(w_t - u)$$

has finite expectation, by the dominated convergence theorem we get

$$\liminf_{n \to \infty} E\left(\rho(w_t - u_n) I_{B_{tD_n}}\right) = E(\rho(w_t - \overline{u})) \geq m_0$$

contradicting (103). This proves (102) for the case (i).

Consider now case (ii). Let D_0, d_1 and k_1 as in the proof of (99). Observe that in B_{tD_0} the condition $|u| \geq k_1$ implies $|w_t - u| \geq d_1$ and therefore

$$\lim_{n \to \infty} E\left(\rho(w_t - u_n) I_{B_{tD_n}}\right) \geq \inf_{|u| \geq k_1} E\left(\rho(w_t - u) I_{B_{tD_0}}\right) \geq 4 m_0 P(B_{tD_0}) \geq 2 m_0$$

contradicting (103). This proves (102) for the case (ii). This completes the proof of the Lemma for the case of unbounded ρ.

Proof of Theorem 4.

Let $\widehat{\gamma}_2$ be as defined in (17). From Lemma 11 we have that

$$\liminf_{T \to \infty} M_{Tk}^{*}(\widehat{\gamma}_{T,2}) > m_0 + \delta \text{ a.s.}$$

for some $\delta > 0$. On the other hand, by Theorem 2 we have that $\widehat{\boldsymbol{\gamma}}_T$ as defined in (14) satisfy $\lim_{T\to\infty} M_T(\widehat{\boldsymbol{\gamma}}_T) = m_0$. This proves the Theorem.

Acknowledgments. This research was partially supported by a grant from the Fundación Antorchas, Argentina, grant X611 from the University of Buenos Aires and grant 03–06277 from the Agencia Nacional de Promoción Científica y Tecnológica, Argentina.

REFERENCES

Amemiya, T. ,1985, Advanced Econometrics. Cambridge, Harvard University Press.

Berkes, I., Horváth, L. and P. Kokoszka , 2003, GARCH processes: structure and estimation. Bernoulli 9, 201-207.

Bianco, A. M., García Ben, M. and V. J. Yohai , 2005, Robust estimation for linear regression with asymmetric errors, Canadian Journal of Statistics, 33.

Bianco, A. M., García Ben, M., Martínez, E. J. and V.J. Yohai , 1996, Robust procedures for regression models with ARIMA Errors', COMPSTAT 96, Proceedings in Computational Statistics, 27–38, Physica-Verlag.

Boente, G. , Fraiman, R. and V. J. Yohai, 1987, Qualitative robustness for general stochastic processes, Annals of Statistics, 15, 1293-1312.

Bollerslev, T. , 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307-27.

Bollerslev, T. , Chou, R. and K. Kroner,1992, ARCH modelling in finance: a review of theory and empirical evidence, Journal of Econometrics, 52, 5-60.

Bougerol, P. and N. Picard, 1992, Stationarity of GARCH processes and of some nonnegative time series, Journal of Econometrics, 52, 115-127.

Boussama, F, 2000, Asymptotic normality for the quasi-maximum likelihood estimator of a GARCH model, Comptes Rendus de l'Académie des Sciences Paris, Serie I, 331, 81-84.

Carnero, A., Peña, D. and E. Ruiz, 2001, Outliers and conditional autoregressive heteroscedasticity in time series, Estadística, 53, 143-213.

Davidson, J. , 1994, Stochastic Limit Theory: An Introduction for Econometricians, New York: Oxford University Press.

Denby, L. and R. D. Martin, 1979, Robust estimation of the first–order autoregressive parameter, Journal of the American Statistical Association, 74, 140–146.

Elie, L. and T. Jeantheau, 1995, Estimation in conditionally heteroskedastic models, Comptes Rendus de l'Académie des Sciences Paris, Serie I, 320, 1255-1258.

Engle, R. F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, Econometrica, 50, 987-1008.

Engle, R. F., Lilien, D. M. and R. P. Robins, 1987, Estimating time-varying risk premia in the term structure: the ARCH-M model, Econometrica, 55, 391-407.

Christian F. and J. M.. Zakoïan, 2004, Maximum likelihood estimation of pure GARCH and

ARMA-GARCH processes, Bernoulli 10, 605-637.

Franses, P. H. and D. van Dijk, 2000, Outlier detection in GARCH models. Research report EI-9926/RV, Econometric Institute, Erasmus University, Rotterdam.

Geweke, J. , 1986, Comment- modeling persistency of conditional variances, Econometric Review, 5, 57-61.

Giraitis, L., Kokoszka, P. and R. Leipus, 2000, Stationary ARCH models: dependence structure and Central Limit Theorem, Econometric Theory, 16, 3-22.

Hampel, F., 1971, A general qualitative definition of robustness, Annals of Mathematical Statistics, 42, 1887-1896.

Hall, P. and Q. Yao, 2003, Inference in ARCH and GARCH models with heavy-tailed errors, Econometrica, 71, 285-317.

Huber, P. J. , 1964, Robust estimation of a location parameter, Annals of Mathematical Statistics, 35, 73-101.

Huber, P. J., 1973, Robust regression: asymptotics, conjectures and Monte Carlo, Annals of Statistics, 1, 799-821.

Huber, P. J., 1981, Robust Statistics, Wiley, N. Y.

Jiang, J., Zhao, Q., and Y.V. Hui, 2001, Robust modelling of ARCH models, Journal of Forecasting, 20, 111-133.

Lee, S. and B. Hansen, 1994, Asymptotic theory for the GARCH(1,1) Quasi-Maximum Likelihood estimator, Econometric Theory, 10, 29-52.

Li, J. and C. Kao, 2002, Bounded influence estimation and outlier detection for GARCH models with and application to foreign exchange rates, Working Paper presented at the 57th European Meeting of the Econometric Society.

Lumsdaine, R.L., 1996, Consistency and asymptotic normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica, 64, 575-596.

Martin, R. D., Samarov, A. and W. Vandaele, 1983, Robust methods for ARIMA models, United States Bureau of Census, Applied Time Series Analysis of Economic Data, 153–177.

Mendes, B.V. M. , 2000, Assessing the bias of maximum likelihood estimates of contaminated GARCH Models, Journal of Statistical Computation and Simulation., 67, 359-376.

Mendes, B.V. M. and A. M. Duarte, 1999, Robust estimation for ARCH models. Revista de Econometria, 19, 138-180.

Muler, N. and V. J. Yohai , 2002, Robust estimates for ARCH Processes, Time Series Analysis, 23 , 341-375.

Nelson, D. , 1990, Stationarity and persistence in the GARCH(1,1), Model. Econometric Theory, 6, 318-334.

Nelson, D. B., 1991, Conditional heteroskedasticity in asset returns: A New Approach, Econometrica, 59, 347-370.

Park, B., 2002, An outlier robust GARCH model and forecasting volatility of exchange rate returns, Journal of Forecasting, 21, 381-393.

Peng, L. and Q. Yao, 2003, Miscellanea. Least absolute deviations estimation for ARCH and GARCH models, Biometrika, 90, 967-975.

Rieder, H. , Ruckdeschel, P. and M. Kohl, 2002, Robust estimation for time series models based on infinitesimal neighborhoods, Working Paper presented at ICORS 2002, Vancouver.

Ronchetti E. and F. Trojani, 2001, Robust inference with GMM estimators, Journal of Econometrics, 101, 37-69.

Sakata, S. and H. White,1998, High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility, Econometrica, 66 (3),529-567.

Straumann, D. and T. Mikosh, 2003, Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equation approach. Technical report, University of Copenhagen.

van Dijk, D., Lucas, A. and P. H. Franses, 1999 Testing for ARCH in the presence of additive outliers. Journal of Applied Econometrics, 14, 539-562

Yohai, V.J., 1987, High breakdown-point and high efficiency robust estimates for regression, Annals of Statistics, 15, 642-656.