Available online at www.sciencedirect.com

ScienceDirect

Journal of Statistical Planning and Inference 138 (2008) 2918–2940

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Robust estimates for GARCH models☆

Nora Muler[a,*], Victor J. Yohai[b]

[a]*Departamento de Matemáticas y Estadística, Universidad Torcuato Di Tella, Miñones 2177, 1428 Buenos Aires, Argentina*
[b]*Universidad de Buenos Aires and CONICET, Argentina*

Received 31 January 2007; received in revised form 18 October 2007; accepted 8 November 2007
Available online 28 November 2007

Abstract

In this paper we present two robust estimates for GARCH 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 proposed estimates compare favorably with respect to other robust estimates. Moreover, we consider some real examples with financial data that illustrate the behavior of these estimates.

Keywords: GARCH models; Robust estimation; M-estimates; Outliers

1. Introduction

In a seminal 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 et al. (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 for the general GARCH(p, q) model among others by Berkes et al. (2003), Straumann and Mikosch (2006) and Christian and Zakoïan (2004). These results prove that if the innovation has four moments, then the QML-estimate is consistent and has asymptotically normal distribution.

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.

☆ This research was partially supported by Grants X-094 from the Universidad de Buenos Aires, PIP 5505 from CONICET and PICT 21407 from ANPCYT, Argentina.
* Corresponding author. Tel./fax: +54 11 5169 7000.
E-mail addresses: nmuler@utdt.edu (N. Muler), vyohai@dm.uba.ar (V.J. Yohai).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.11.003

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 modelled 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. Mendes and Duarte (1999) defined a class of constrained M-estimates and Muler and Yohai (2002) introduced a class of estimates based on a τ-scale estimate combined with robust filtering. Jiang et al. (2001) proposed L_1 estimates of modified ARCH models. Franses and van Dijk (2000) and Carnero et al. (2001) used diagnostic procedure for detecting outliers in GARCH models. Rieder et al. (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 (MSE) on infinitesimal neighborhoods of contamination. Robust tests for ARCH heteroskedasticity were proposed by van Dijk et al. (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 and of the absolute values of the preceding observations. The proposed estimator is based on a least absolute deviation (LAD) criterion. Peng and Yao (2003) proposed estimates for the GARCH model which are also variations of the LAD criterion. Sakata and White (1998) proposed S-estimates for the GARCH model. 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 Student with a small degrees 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.

None of the robust estimates mentioned above for the GARCH model satisfies (H2). The reason is that they are based on predictors of the conditional variance which are very sensitive to outliers. In fact, the presence of a large outlier in one period may influence the predictors of the conditional variance for several subsequent periods. A Monte Carlo study, reported in Section 4, confirms that a small fraction of additive outliers may have a large influence on these estimates, and therefore they do not satisfy (H2).

In reference to (H1), in Table 1 of Section 4 we show the asymptotic efficiencies of most of these estimates. These efficiencies, except for the maximum likelihood estimate based on a Student distribution, are rather low.

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), because they also depend on the predictors of the conditional variance which are very sensitive to outliers.

To improve robustness, we propose a modification of the M-estimates which includes an additional mechanism that bounds the propagation of the effect of one outlier on the subsequent predictors of the conditional variances. 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.

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 where the proposed estimates are compared with the QML-estimate and other robust estimates. 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 several estimates. Section 6 contains some concluding remarks. An Appendix contains some of the proofs. For brevity sake we omit several proofs which can be found in a technical report by Muler and Yohai (2007).

2. Robust estimates for GARCH(p, q) models

A series $x_1, \ldots, x_T$ is a centered GARCH (p, q) process if $x_t = \sigma_t z_t$, where $z_1, z_2, \ldots, z_T$ are i.i.d. random variables with a continuous density f such that $E(z_t) = 0$ and $\mathrm{var}(z_t) = 1$ ($\mathrm{var}(x)$ 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 $\alpha_i \geqslant 0$, $1 \leqslant i \leqslant p$, $\beta_i \geqslant 0$, $1 \leqslant i \leqslant q$ and $\alpha_0 > 0$. We denote $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)$, $\beta = (\beta_1, \ldots, \beta_q)$ and $\gamma = (\alpha, \beta)$. 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

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

see for example Bollerslev (1986) and Bougerol and Picard (1992). In this case

$$\mathrm{var}(x_t) = \frac{\alpha_0}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i)}. \tag{2}$$

The following condition is required for identification of the GARCH parameters

$$A(x) = \sum_{i=1}^{p} \alpha_i x^i \quad \text{and} \quad B(x) = 1 - \sum_{i=1}^{q} \beta_i x^i \quad \text{are coprimes.} \tag{3}$$

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

$$\sigma_t^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \beta_i} + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} \beta_{j_1} \cdots \beta_{j_k} x_{t-i-j_1-\cdots-j_k}^2. \tag{4}$$

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

$$g(w) = f(\mathrm{e}^{w/2})\mathrm{e}^{w/2}. \tag{5}$$

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

$$g_0(w) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(1/2)(\mathrm{e}^w - w)}. \tag{6}$$

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

$$h_t(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{q} b_i h_{t-i}(\mathbf{c}), \tag{7}$$

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

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

for all $t \leqslant 0$. These initial conditions are the same as those used by Hall and Yao (2003).

From (4) we obtain that

$$h_t(\mathbf{c}) = \frac{a_0}{1-\sum_{i=1}^{q} b_i} + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x_{t-i-j_1-\cdots-j_k}^2 \; I_{\{t-i-j_1-\cdots j_k \geqslant 1\}},$$

where I_A denotes the indicator function of A.

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

$$-\frac{1}{2}\sum_{t=p+1}^{T} \frac{x_t^2}{h_t(\mathbf{c})} - \frac{1}{2}\sum_{t=p+1}^{T} \log h_t(\mathbf{c}) \tag{8}$$

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

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

Then, the QML-estimate can be obtained by maximizing

$$L_{0,T}(\mathbf{c}) = \sum_{t=p+1}^{T} \log(g_0(y_t - \log h_t(\mathbf{c}))), \tag{9}$$

where the function g_0 is given by (6). Therefore, maximizing (8) is equivalent to minimizing

$$M_{0,T}(\mathbf{c}) = \frac{1}{T-p}\sum_{t=p+1}^{T} \rho_0(y_t - \log h_t(\mathbf{c})), \tag{10}$$

where

$$\rho_0 = -\log(g_0), \tag{11}$$

and where g_0 is given in (6).

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

In a similar way, it can be proved that the maximum likelihood 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 (5) 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

$$M_T(\mathbf{c}) = \frac{1}{T-p}\sum_{t=p+1}^{T} \rho(y_t - \log h_t(\mathbf{c})). \tag{12}$$

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

$$\widehat{\gamma}_T = \arg\min_{\mathbf{c}\in C} M_T(\mathbf{c}) \tag{13}$$

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 by Huber (1973) for regression.

Define $J(u) = E(\rho(w_t - u))$. The following lemma proved in Muler and Yohai (2007) shows that $J(u)$ is well defined when ρ' is finite.

Lemma 1. *Consider a stationary GARCH model x_t. 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 et al. (1992) proposed using the maximum likelihood 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 and (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 maximum likelihood 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 $\rho(u) = 2\ln(1+e^u) - u/2$ and $\rho'(u) = (3e^u - 1)/(2(1+e^u))$. For the Peng and Yao estimate we have $\rho(u) = |u|$ and $\rho'(u) = \mathrm{sign}(u)$.

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 (7), 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 Martin et al. (1983) and Bianco et al. (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

$$h^*_{t,k}(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i h^*_{t-i,k}(\mathbf{c}) r_k\left(\frac{x^2_{t-i}}{h^*_{t-i,k}(\mathbf{c})}\right) + \sum_{i=1}^{q} b_i h^*_{t-i,k}(\mathbf{c}), \tag{14}$$

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

$$r_k(u) = \begin{cases} u & \text{if } u \leqslant k, \\ k & \text{if } u > k. \end{cases} \tag{15}$$

Note 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 (14) would fit better than the M-estimate corresponding to the GARCH model. This suggests modifying the M-estimate as follows. Let $\widehat{\gamma}_{1,T}$ be defined as in (13) and $\widehat{\gamma}_{2,T}$ by

$$\widehat{\gamma}_{2,T} = \arg\min_{\mathbf{c}\in C} M^*_{Tk}(\mathbf{c}), \tag{16}$$

where

$$M^*_{Tk}(\mathbf{c}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log h^*_{t,k}(\mathbf{c})). \tag{17}$$

If the process is a perfectly observed GARCH process without outliers, then the conditional variances are given by (7). Then the estimate $\widehat{\gamma}_{1,T}$ using these conditional variances generally behaves better than $\widehat{\gamma}_{2,T}$. In this case $\widehat{\gamma}_{2,T}$ is asymptotically biased and we may expect $M_T(\widehat{\gamma}_{1,T}) < M^*_{Tk}(\widehat{\gamma}_{2,T})$. Theorem 4 of Section 3 proves that this holds asymptotically with probability one. As explained above, when there are outliers, $\widehat{\gamma}_{2,T}$ may be preferable and we may expect $M_T(\widehat{\gamma}_{1,T}) > M^*_{Tk}(\widehat{\gamma}_{2,T})$. Then we define the BM-estimate by

$$\widehat{\gamma}^B_T = \begin{cases} \widehat{\gamma}_{1,T} & \text{if } M_T(\widehat{\gamma}_{1,T}) \leqslant M^*_{Tk}(\widehat{\gamma}_{2,T}), \\ \widehat{\gamma}_{2,T} & \text{if } M_T(\widehat{\gamma}_{1,T}) > M^*_{Tk}(\widehat{\gamma}_{2,T}). \end{cases} \tag{18}$$

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 (5) and ρ is a nondecreasing 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 1–3 we prove the consistency and asymptotic normality for any M-estimate defined in (13) as long as (P1) holds and ρ' is bounded. In Theorems 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 $\mathbf{X}_t = (\ldots, x_{t-1}, x_t)$ corresponding to a GARCH(p, q) process up to time t with parameter $\gamma = (\alpha, \beta)$, 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 $\gamma = \mathbf{c}$. Then the following recursive relationship is satisfied:

$$\widetilde{h}_t(\mathbf{c}) = a_0 + \sum_{i=1}^{p} a_i x^2_{t-i} + \sum_{i=1}^{q} b_i \widetilde{h}_{t-i}(\mathbf{c}). \tag{19}$$

Denote by $R^n_+ = \{\mathbf{x} = (x_1, \ldots, x_n)\colon x_i \geqslant 0, 1 \leqslant i \leqslant n\}$. The following theorem shows the Fisher consistency of the M-estimates of the GARCH model and gives a sufficient condition for (P1).

Theorem 1. *Let x_t be a stationary GARCH(p, q) process satisfying* (1). *Let $y_t = \log(x^2_t)$ 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* (3) *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 $a_i = e^{u_0}\alpha_i$, $0 \leqslant i \leqslant p$, $b_i = \beta_i$, $1 \leqslant i \leqslant q$.*
(ii) *Assume that $w_t = \log(z^2_t)$ 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 $u_0 = 0$.*

Remark 1. According to part (i) of Theorem 1, the M-estimate of α 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

$$C_\delta = \left\{ (\mathbf{a}, \mathbf{b}) : \mathbf{a} \in R_+^{p+1}, \mathbf{b} \in R_+^q, a_0 \in [\delta, 1/\delta],\ \sum_{i=1}^p a_i \geqslant \delta, \sum_{i=1}^p a_i + \sum_{j=1}^q b_j \leqslant 1-\delta \right\}. \tag{20}$$

The set C in (13) and (16) is taken as C_{δ_0} for some $\delta_0 > 0$.

Put $\gamma = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. The following two theorems state the consistency and asymptotic normality of the M-estimates of γ.

Theorem 2. *Suppose that all the assumptions of Theorem* 1 *hold. Let* $\widehat{\gamma}_T$ *be defined as in* (13) *with* $C = C_{\delta_0}$ *given by* (20). *We also assume that* (P1) *is satisfied with* $u_0 = 0$, *that* ρ *has a bounded derivative, and that* $\gamma \in C$. *Then* $\widehat{\gamma}_T \to \gamma$ *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* $T^{1/2}(\widehat{\gamma}_T - \gamma)$ *converges in distribution to a* $N(\mathbf{0}, V)$ *where*

$$V = a(\psi, g)\left(E_g\left(\frac{1}{\widetilde{h}_t^2(\gamma)}\nabla\widetilde{h}_t(\gamma)\nabla\widetilde{h}_t(\gamma)'\right)\right)^{-1}, \tag{21}$$

where

$$a(\psi, g) = \frac{E_g(\psi^2(w_t))}{E_g^2(\psi'(w_t))}. \tag{22}$$

and where ∇h *denotes gradient of* h.

Remark 2. For the case of the QML-estimate, we have $\rho = \rho_0$ given in (11). Then the assumption that ρ' is 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 (21). See Berkes et al. (2003). Then the relative asymptotic efficiency of the M-estimate with respect to the QML-estimate is given by

$$\mathrm{AEF} = \frac{a(\psi_0, g)}{a(\psi, g)}, \tag{23}$$

where $\psi_0 = \rho_0'$. 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.

Remark 3. We use the differentiability conditions in Theorem 3 to simplify the proof. However, we conjecture that these conditions may be weakened. Note also that from the practical point of view these assumptions are not very restrictive, since any Lipschitz piecewise differentiable function like Huber's ρ function can be approximated by another function satisfying our assumptions. In Remark 4 we give an example where ρ' is not differentiable and Theorem 3 holds.

Remark 4. Integrating by parts $\int_{-\infty}^{\infty} \psi'(u)g(u)\,\mathrm{d}u$ we get

$$E_g(\psi'(w_t)) = -\int_{-\infty}^{\infty} \psi(u)g'(u)\,\mathrm{d}u$$

and therefore we have

$$a(\psi, g) = \frac{E_g(\psi^2(w_t))}{\left(\int_{-\infty}^{\infty} \psi(u)g'(u)\,\mathrm{d}u\right)^2}. \tag{24}$$

Peng and Yao (2003) show that the LAD estimate which corresponds to $\rho(u) = |u - u_0|$, where u_0 minimizes $J(u) = E(|w_t - u|)$, is also asymptotically normal with covariance matrix given by (21) with $a(\psi, g)$ given by (24). Note that in this case $\psi(u) = \mathrm{sign}(u - u_0)$ and so $\int_{-\infty}^{\infty} \psi(u)g'(u)\,\mathrm{d}u = -2g(u_0)$.

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* $\lim_{|u|\to\infty}\rho(u)=\sup_u \rho(u)$. *Let* $\widehat{\gamma}_T$ *be the M-estimate defined by* (13) *and* $\widehat{\gamma}_T^B$ *the BM-estimate defined by* (18), *then* $\lim_{T\to\infty} P(\widehat{\gamma}_T^B=\widehat{\gamma}_T)=1$.

Remark 5. 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 nondecreasing.

From Theorems 2–4 we derive the following result.

Theorem 5. *Assume that the distribution of* z_t *gives positive probability to the complement of any compact and that* $\lim_{|u|\to\infty}\rho(u)=\sup_u \rho(u)$. *Then, Theorems* 2 *and* 3 *also hold for the BM-estimate* $\widehat{\gamma}_T^B$.

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 S-estimate proposed by Sakata and White (1998) using a M-scale estimate with 50% breakdown point (S), (v) the M-estimate based on a loss function $\rho_1=m_1(\rho_0)$, where ρ_0 is given in (11) and m_1 is a nondecreasing, bounded and close to the identity function which is defined below (M_1), (vi) A BM-estimate as defined in (18) with $\rho=\rho_1$ and $k=5.02$ (BM_1), (vii) 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 (viii) A BM-estimate as defined in (18) with $\rho=\rho_2$ and $k=3$ (BM_2). As we will see below, M_2 and BM_2 are more robust but less efficient than M_1 and BM_1, respectively.

The function m_1 is a smoothed version of

$$m(x)=\begin{cases} x & \text{if } x\leqslant 4,\\ 4 & \text{if } x>4\end{cases}$$

which coincides with a fourth order polynomial in the interval [4, 4.30]. The exact formula for m_1 is given by

$$m_1(x)=\begin{cases} x & \text{if } x\leqslant 4,\\ P(x) & \text{if } 4<x\leqslant 4.3,\\ 4.15 & \text{if } x>4.3,\end{cases}$$

where

$$P(x)=\frac{2}{(b-a)^3}\left(\frac{1}{4}(x^4-a^4)-\frac{1}{3}(2a+b)(x^3-a^3)+\frac{1}{2}(a^2+2ab)(x^2-a^2)\right) - \frac{2a^2b}{(b-a)^3}(x-a)-\frac{1}{3(b-a)^2}(x-a)^3+x$$

and where $a=4$ and $b=4.3$. The polynomial $P\,(x)$ was obtained using the following constraints:

$$P(a)=a,\quad P'(a)=1,\quad P'(b)=P''(a)=P''(b)=0$$

which makes $m_1(x)$ continuous and smooth. This function is shown in Fig. 1.

Note that 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 make estimates M_1 and BM_1 more efficient than estimates M_2 and BM_2 (see Table 1). As a counterpart we see in our Monte Carlo results that the first estimates are going to be less robust than the second ones. To asses how close to ρ_0 are ρ_1 and ρ_2, we note that

$$P(\rho_1(w_t)=\rho_0(w_t))=0.96$$

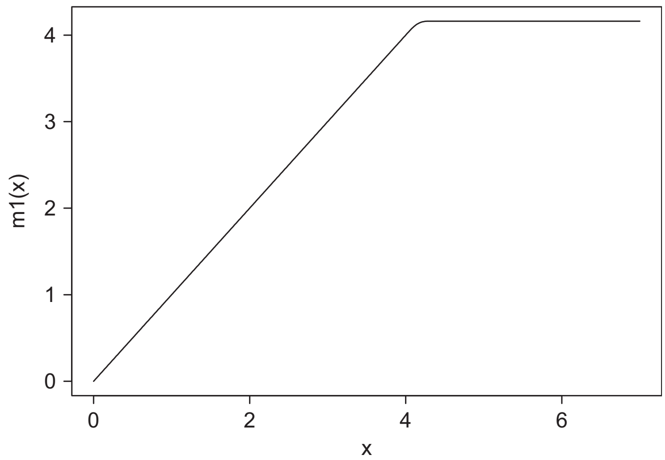

Fig. 1. Function m_1.

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

Estimate	QML	SML	LAD	S	M_1 and BM_1	M_2 and BM_2
EFF	1	0.79	0.37	0.33	0.83	0.67

Table 2
Mean square errors (MSEs) 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
S	0.12	0.28	0.014	0.29	0.009	0.33
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

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 (15) for BM_1 was taken as equal to 5.02. In order to asses how close is $r_{5.02}$ to the identity we note that $P(r_{5.02}(z_t^2) = z_t^2) = 0.975$, when z_t is N(0,1). For the estimate BM_2, in order to gain robustness, we chose $k = 2.72$. This value is such $P(r_{2.72}(z_t^2) = z_t^2) = 0.90$. 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 than BM_1 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 estimate SML, -0.787 for the estimate LAD and -0.306 for the estimate S. 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

Table 3
Mean square errors for a normal GARCH(1,1) model with 5% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$

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
S	0.38	0.049	0.026	0.65	0.086	0.059
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

Outlier size: $d\sigma_t$.

Table 4
Mean square errors for a normal GARCH(1,1) model with 10% of additive outliers and parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.4$

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
S	0.50	0.14	0.03	0.53	0.20	0.048
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

Outlier size: $d\sigma_t$.

Table 5
Mean square errors for a Student GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0$

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
S	0.110	0.024	0.016
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

efficiencies of estimates M_1, BM_1 and SML are quite high, the asymptotic efficiencies of estimates M_2 and BM_2 are intermediate and the asymptotic efficiencies of estimates LAD and S are quite low.

We report here the results using a GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$. Similar results obtained for the ARCH(2) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\alpha_2 = 0.4$ can be found in Muler and Yohai (2007). Other GARCH(1,1) and ARCH(2) models were also 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 (20) was taken as equal to 0.01.

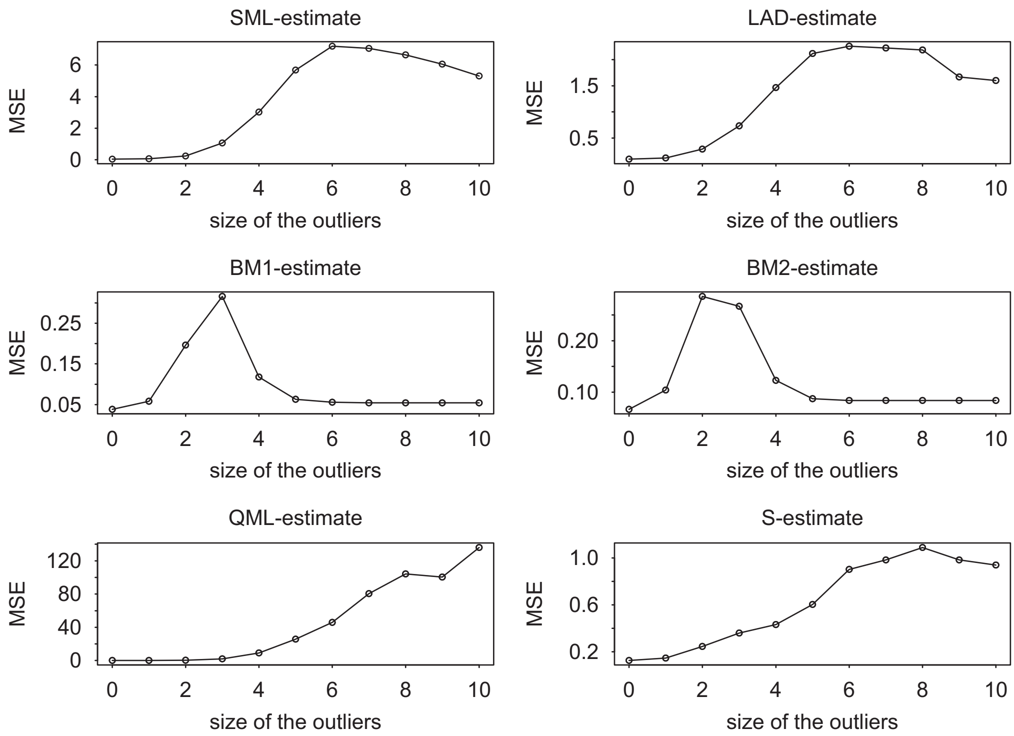

Fig. 2. Mean square errors of α_0 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination.

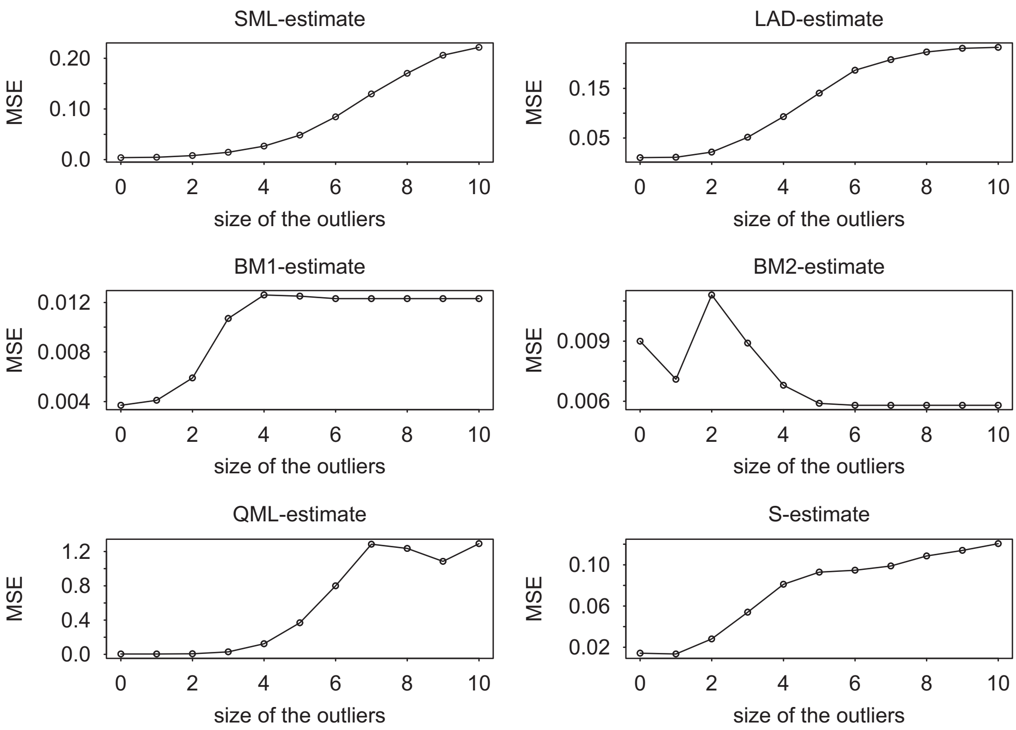

Fig. 3. Mean square errors of α_1 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0 = 1$, $\alpha_1 = 0.5$ and $\beta_1 = 0.4$ and 5% of additive outlier contamination.

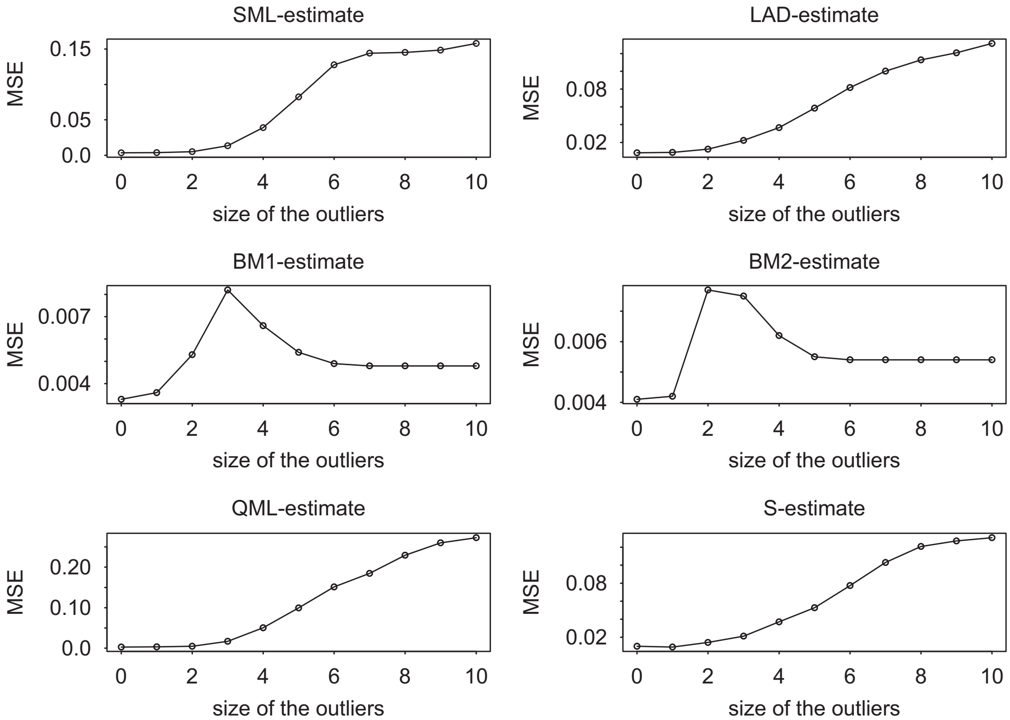

Fig. 4. Mean square errors of β_1 as a function of the outlier size d for the normal GARCH(1,1) model with parameters $\alpha_0=1$, $\alpha_1=0.5$ and $\beta_1=0.4$ and 5% of additive outlier contamination.

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 three degrees of freedom.

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

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

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

Table 2 shows the mean square errors (MSEs) for the normal GARCH(1,1) model in the case of no outliers. In this table we show the efficiency (EFF) of the estimates with respect to the estimate QML. We observe that in this case estimates M_1 and BM_1 behave similarly. The same happens with M_2 and BM_2.

In Table 3 we show the MSE for 5% contaminated samples for the normal GARCH(1,1) model. Note that the estimate QML can be seriously affected by outliers, especially for $d=5$. Although estimates LAD, SML, S, M_1 and M_2 are not so much affected by outliers we can see that BM_1 and BM_2, in general, behave much better.

In Table 4 we report the MSE for the normal GARCH(1,1) 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 estimate BM_2 behaves better than the others.

Table 5 shows the MSE for a GARCH(1,1) model where z_t has a Student distribution with three degrees of freedom. As may be expected the smallest MSE corresponds to the estimate SML. The other robust estimates behave quite similarly and better than the QML.

In Figs. 2–4 we plot the MSEs as a function of the outlier size d for estimates QML, SML, LAD, S, 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 observe that estimates BM_1 and BM_2 behave more robustly than the others.

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. This plot suggests 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 estimates QML, SML, LAD, S and BM_1 defined as in Section 4. In Table 6 we show the values of these estimates for the four series. We observe important differences among the estimates which may be due to the presence of outliers.

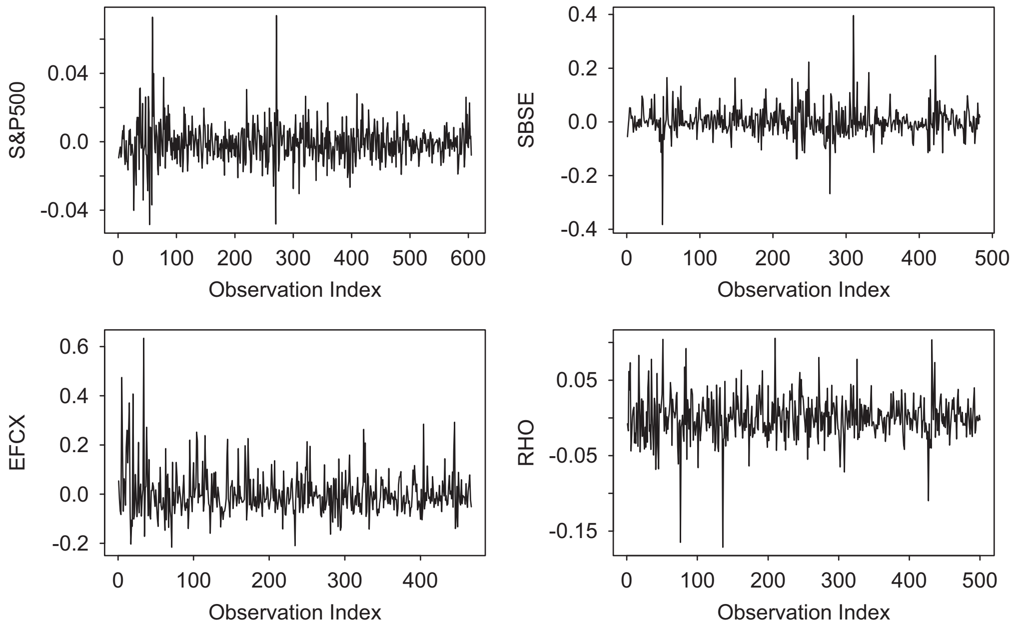

Fig. 5. Plot of the daily return series.

Table 6
Fitted GARCH(1,1) models for the daily returns series

Estimates	S&P 500 $\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	SBSE $\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$
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
S	5.0×10^{-6}	0.10	0.78	1.54×10^{-4}	0.12	0.64
BME_1	4.5×10^{-6}	0.10	0.84	4.1×10^{-4}	0.26	0.50
	EFCX			RHO		
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
S	2.1×10^{-3}	0.12	0.13	1.9×10^{-4}	0.097	0.52
BME_1	2.6×10^{-3}	0.259	0.20	2.4×10^{-4}	0.306	0.30

Table 7
Truncated variance (σ^2_{TR}) and rank correlation (τ) for the daily returns series

Estimates	S&P 500		SBSE		EFCX		RHO	
	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
S	1.32	0.018	1.52	0.039	1.52	−0.008	1.091	0.036
BME_1	0.98	0.022	1.02	−0.015	0.96	−0.043	1.045	−0.035

Let x_t, $1 \leqslant t \leqslant T$, 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 estimates QML, SML, LAD and S, $\widehat{\sigma}_t^2$ is recursively computed by

$$\widehat{\sigma}_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 x_{t-1}^2 + \widehat{\beta}_1 \widehat{\sigma}_{t-1}^2, \quad 2 \leqslant t \leqslant T. \tag{25}$$

In the case of the BM_1, $\widehat{\sigma}_t^2$ is also given by (25) if $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ coincides with $\widehat{\gamma}_1$ defined in (13) or by

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

if $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}_1)$ coincides with $\widehat{\gamma}_{2,T}$ defined in (16).

When x_t follows a GARCH model, the series z_t have the following two properties (a) $\mathrm{var}(z_t) = 1$ 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

$$\widehat{z}_t = \frac{x_t}{\widehat{\sigma}_t}, \quad 2 \leqslant t \leqslant T. \tag{26}$$

If the estimate used 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

$$\sigma^2_{\mathrm{TR}} = \frac{1.605}{T_1} \sum_{t=1}^{T_1} \widehat{z}^2_{(t)}, \tag{27}$$

where $\widehat{z}^2_{(1)} \leqslant \cdots \leqslant \widehat{z}^2_{(T-1)}$ are the order statistics of $(\widehat{z}_2^2, \ldots, \widehat{z}_T^2)$, 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 $\widehat{z}_{t-1}^2$ and $\widehat{z}_t^2$, which is a robust correlation measure. We denote this estimate by τ.

Table 7 shows the value of σ^2_{TR} and τ corresponding to estimates QML, SML, LAD, BM_1 and S for the four series.

We observe that the value of σ^2_{TR} for the estimate 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 estimate BM_1 performs better than the others for series S&P 500, SBSE. Instead, for the series EFCX the estimate SML seems preferable. For the RHO series estimates BM_1 and S have a similar behavior and perform better than the others.

6. Concluding remarks

In this paper we present two classes of robust estimates for GARCH models: M- and BM-estimates. The BM-estimates include a mechanism that restricts the propagation of the outlier effect in such a way that the influence of past variances on the present observation is bounded. A Monte Carlo study shows that for the GARCH(1,1) model, the QML-estimate may practically collapse when there is 5% of outlier contamination. All the robust estimates considered in the

Monte Carlo study are in general less influenced by outliers. However, under outlier contamination, the BM-estimates generally behave much better than the M-estimate and the other robust methods.

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

Appendix

Proof of Theorem 1. (i) Let $\gamma=(\alpha,\beta)$ be the true parameter. Then, we can write

$$M(\mathbf{c})=E\left(\rho\left(w_t-\log\left(\frac{\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\gamma)}\right)\right)\right), \tag{28}$$

where $w_t=\log(y_t/\widetilde{h}_t(\gamma))=\log(z_t^2)$ are i.i.d. random variables with density 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

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

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

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

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

Given a vector $\mathbf{c}=(c_1,\ldots,c_k)\in R^k$, we denote by $\|\mathbf{c}\|=(\sum_{i=1}^k c_i^2)^{1/2}$ and given a $k\times k$ matrix A, $\|A\|=\sup_{\mathbf{c}\in R^k}\|A\mathbf{c}\|/\|\mathbf{c}\|$. Given a differentiable function $g(x_1,\ldots,x_k)$, we denote by $\nabla g=(\partial g/\partial x_1,\ldots,\partial g/\partial x_k)$ and by $\nabla^2 g$ the $k\times k$ matrix whose (i,j) element is $\partial^2 g/\partial x_i\partial x_j$. $\square$

The following Lemmas 2–4 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). *Let $h_t(\mathbf{c})$ be as defined in* (7) *and $\widetilde{h}_t(\mathbf{c})$ as defined in* (19). *Then*:

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

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

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

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

for all $t\geqslant p+1$.

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

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

for all $t\geqslant p+1$.

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

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

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

for all $t\geqslant p+1$.

Proof. (i) Hall and Yao (2003) show that

$$\widetilde{h}_t(\mathbf{c}) = \frac{a_0}{1-\sum_{i=1}^{q} b_i} + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x_{t-i-j_1-\cdots-j_k}^2, \tag{29}$$

and then

$$h_t(\mathbf{c}) = \frac{a_0}{1-\sum_{i=1}^{q} b_i} + \sum_{i=1}^{p} a_i x_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_i=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} x_{t-i-j_1-\cdots-j_k}^2 \; I_{t-i-j_1-\cdots j_k \geqslant 1} \tag{30}$$

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

Then, from (29) and (30) we obtain

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

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

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

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

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

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

Note 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) and (v) in Peng and Yao (2003). The remaining points are proved in Muler and Yohai (2007). □

Lemma 3. *Let C be as in* (20) *and*

$$\widetilde{M}_T(\mathbf{c}) = \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log \widetilde{h}_t(\mathbf{c})). \tag{31}$$

Then under the assumptions of Theorem 2, *we have*

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

Proof. We start proving

$$E\left(\sup_{\mathbf{c}\in C}(|\rho(y_t - \log \widetilde{h}_t(\mathbf{c}))|)\right) < \infty. \tag{32}$$

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

$$E\left(\sup_{\mathbf{c}\in C}(|y_t - \log \widetilde{h}_t(\mathbf{c})|)\right) < \infty. \tag{33}$$

We also have

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

Since by Lemma 1 (a) $E(|w_t|) < \infty$, to prove (32) it is enough to show that

$$E\left(\sup_{\mathbf{c}\in C}(|\log \widetilde{h}_t(\mathbf{c})|)\right) < \infty. \tag{34}$$

From (20) and (29) we have that

$$E\left(\sup_{\mathbf{c}\in C}\widetilde{h}_t(\mathbf{c})\right) < \infty, \tag{35}$$

and from (20) we obtain $\delta_0 \leqslant \inf_{\mathbf{c}\in C}\widetilde{h}_t(\mathbf{c})$. Then (34) follows from (35) and the lemma follows from Lemma 3 of Muler and Yohai (2002). □

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

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

a.s.

Proof. We have

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

Let $K = \sup|\rho'| < \infty$. Then, from Lemma 2(i) and (20) we have that there exists $0 < \vartheta < 1$, and a finite positive random variable W such that

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

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

$$A \subset \{\widehat{\gamma}_T \to \gamma\}. \tag{36}$$

Assume that (36) is not true. Then we can find in A a subsequence $\widehat{\gamma}_{T_i}$ such that

$$\widehat{\gamma}_{T_i} \to \overline{\gamma} \tag{37}$$

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

$$M(\mathbf{c}) > M(\gamma) + \varepsilon. \tag{38}$$

From (37) there exists i_0 large enough such that for all $i \geqslant i_0$ we obtain

$$\widehat{\gamma}_{T_i} \in U(\overline{\gamma}), \quad \sup_{\mathbf{c}\in C}|M_{T_i}(\mathbf{c}) - M(\mathbf{c})| < \frac{\varepsilon}{2}. \tag{39}$$

Therefore from (38) and (39) for all $i \geqslant i_0$ we obtain

$$M_{T_i}(\widehat{\gamma}_{T_i}) = M_{T_i}(\widehat{\gamma}_{T_i}) - M(\widehat{\gamma}_{T_i}) + M(\widehat{\gamma}_{T_i}) > M(\gamma) + \frac{\varepsilon}{2}. \tag{40}$$

Using the definition of $\widehat{\gamma}_T$ and (39), we get $M_{T_i}(\widehat{\gamma}_{T_i}) \leqslant M_{T_i}(\gamma) < M(\gamma) + \varepsilon/2$, for all $i \geqslant i_0$. This contradicts (40) 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,*

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

where

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

Proof. From Lemma 2(ii) D_0 is finite, and from (3) 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 $\mathbf{b}'\psi(w_t)\nabla\log(\widetilde{h}_t(\gamma))$ 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 Davidson, 1994, Theorem 24.4) we obtain

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

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* γ *such that*

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

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

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

$$\begin{aligned}&\nabla^2\rho(y_t-\log(\widetilde{h}_t(\mathbf{c})))\\ &\quad=\psi'(y_t-\log\widetilde{h}_t(\mathbf{c}))+\psi(y_t-\log(\widetilde{h}_t(\mathbf{c})))\nabla\log(\widetilde{h}_t(\mathbf{c}))\nabla\log(\widetilde{h}_t(\mathbf{c}))'-\psi(y_t-\log(\widetilde{h}_t(\mathbf{c})))\frac{\nabla^2\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})}.\end{aligned} \tag{41}$$

From (20) and Lemma 2(v) there exists a neighborhood U of γ such that

$$E\left(\sup_{\mathbf{c}\in U}\left\|\frac{\nabla^2\widetilde{h}_t(\mathbf{c})}{\widetilde{h}_t(\mathbf{c})}\right\|\right)<\infty. \tag{42}$$

Then by Lemma 2(ii), (41), (42) and the fact that ψ and ψ' are continuous and bounded we get

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

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

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

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

and

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

Then $A(\gamma)=E(\psi'(w_t))D_0$. 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 $\widetilde{h}_t(\gamma)$'s instead of the $h_t(\gamma)$'s. The proofs can be found in Muler and Yohai (2007).

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

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

a.s.

Lemma 8. *Suppose that all the assumptions of Theorem* 3 *hold. Then, there exists a neighborhood U of γ such that*

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

a.s.

Proof of Theorem 3. From Lemmas 5 and 7 we have

$$\frac{1}{\sqrt{T-p}}\sum_{t=p+1}^{T}\nabla\rho(y_t-\log(h_t(\gamma)))\to_D N(\mathbf{0},E(\psi^2(w_t))D_0),\tag{43}$$

and from Lemmas 6(i) and 8 we get that there exists a neighborhood U of γ such that

$$\lim_{T\to\infty}\sup_{\mathbf{c}\in U}\left\|\frac{1}{T-p}\sum_{t=p+1}^{T}\nabla^2\rho(y_t-\log(h_t(\mathbf{c})))-A(\mathbf{c})\right\|=0\quad\text{a.s.}\tag{44}$$

From (43), (44) and Theorem 2 we get that

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

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

The following Lemmas 9–11 are going to be used to prove Theorem 4.

Lemma 9. *Let $\mathbf{Y}_t$ be an ergodic process in R^m and $g:R^m\times R\to R$, a continuous function satisfying*: (i) *There exists $g_0:R^m\to R$ such that $|g(\mathbf{Y}_t,u)|\leqslant g_0(\mathbf{Y}_t)$, and $g_0(\mathbf{Y}_t)$ integrable and* (ii) $\lim_{u\to\infty}g(\mathbf{Y}_t,u)=g^+(\mathbf{Y}_t)$ *a.s. and* $\lim_{u\to-\infty}g(\mathbf{Y}_t,u)=g^-(\mathbf{Y}_t)$ *a.s. Then*

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

Proof. From Muler and Yohai (2002) we have

$$\lim_{T\to\infty} \sup_{u\in K} \left|\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E(g(\mathbf{Y}_t, u))\right| = 0 \quad \text{a.s.} \tag{45}$$

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

$$\lim_{T\to\infty} \sup\nolimits_{u\geqslant\overline{u}} \left|\frac{1}{T}\sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E(g(\mathbf{Y}_t, u))\right| \leqslant \varepsilon \quad \text{a.s.} \tag{46}$$

and

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

Since both proofs are similar we only show (46). To this purpose, it is enough to prove that

$$\lim_{T\to\infty} \sup_{u\geqslant\overline{u}} \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E(g(\mathbf{Y}_t, u)) \leqslant \varepsilon \quad \text{a.s.} \tag{47}$$

and

$$\lim_{T\to\infty} \inf_{u\geqslant\overline{u}} \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{Y}_t, u) - E(g(\mathbf{Y}_t, u)) \geqslant -\varepsilon \quad \text{a.s.} \tag{48}$$

Since the proofs of (47) and (48) are similar, we only show (47).

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

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

and this proves (47). □

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

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

where $h^*_{t,k}(\mathbf{c})$ *is defined in* (14) *and* R_t *is a positive-valued ergodic process.*

Proof. Define $R_t = \sup_{\mathbf{c}\in C} \widetilde{h}_t(\mathbf{c})$. Then, from (20) and (29) we have that R_t is a positive-valued ergodic processes and from (7) we get

$$\sup_{\mathbf{c}\in C} h_t(\mathbf{c}) \leqslant R_t. \tag{49}$$

We prove by induction on t that

$$h^*_{t,k}(\mathbf{c}) \leqslant h_t(\mathbf{c}). \tag{50}$$

From (14) it follows immediately that $h^*_{t,k}(\mathbf{c}) = h_t(\mathbf{c})$ for all $t \leqslant 0$, Assume now $h^*_{j,k}(\mathbf{c}) \leqslant h_j(\mathbf{c})$, $j \leqslant t$. Then from (14) we have

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

and thereby (50) follows. Then, the lemma follows from (49). □

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*

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

where M^*_{Tk} *is given in* (17).

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

$$\widetilde{h}_t(\gamma) \geqslant \alpha_{i_0} x^2_{t-i_0} = \alpha_{i_0} z^2_{t-i_0} \widetilde{h}_{t-i_0}(\gamma). \tag{51}$$

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 $t \geqslant 1$

$$h^*_{t,k}(\mathbf{c}) \leqslant a_0 + \sum_{i=1}^{s} (a_i k + b_i) h^*_{t-i,k}(\mathbf{c}) \leqslant (2+k) \sum_{i=1}^{s} h^*_{t-i,k}(\mathbf{c}) \tag{52}$$

and

$$\sum_{i=1}^{s} h^*_{t-i,k}(\mathbf{c}) \leqslant (2+k) \sum_{i=1}^{s} h^*_{t-i-1,k}(\mathbf{c}) + \sum_{i=2}^{s} h^*_{t-i,k}(\mathbf{c}) \leqslant 2(2+k) \sum_{i=1}^{s} h^*_{t-i-1,k}(\mathbf{c}). \tag{53}$$

By Lemma 10, there exists a positive-valued ergodic process R_t such that $\sup_{\mathbf{c}\in C} h^*_{t,k}(\mathbf{c}) \leqslant R_t$. Then, from (52) and (53) we have that

$$h^*_{t,k}(\mathbf{c}) \leqslant 2^{i_0} (2+k)^{i_0+1} \sum_{i=i_0+1}^{s+i_0} R_{t-i}. \tag{54}$$

Let us define the ergodic processes

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

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

$$P(N_t > \eta) \geqslant v. \tag{55}$$

Using that $\lim_{u\to\infty} \rho(u) = \sup_u \rho(u) > m_0$, there exists $k_1 > 0$ such that

$$\inf_{u \geqslant k_1} E(\rho(w_t + u)) > m_0. \tag{56}$$

Let us define K as

$$K = \frac{2^{i_0} (2+k)^{i_0+1} \mathrm{e}^{k_1}}{\eta \alpha_{i_0}} \tag{57}$$

and define

$$A_t = \{N_t > \eta,\ z^2_{t-i_0} > K\}. \tag{58}$$

Since $z^2_{t-i_0}$ is independent of N_t, using (55) and the fact that z_t is unbounded, we have

$$a = P(A_t) = vP(z^2_{t-i_0} > K) > 0. \tag{59}$$

From (54), (51) and the choice of K in the definition of A_t we have

$$\sup_{\mathbf{c}\in C} (\log \widetilde{h}_t(\gamma) - \log h^*_{t,k}(\mathbf{c})) \geqslant k_1. \tag{60}$$

We have from (56) that there exists $\delta > 0$ such that

$$(1-a)E(\rho(w_t)) + a \inf_{u \geqslant k_1} E(\rho(w_t+u)) \geqslant m_0 + \delta. \tag{61}$$

Put

$$g(u) = E(\rho(w_t+u))$$

and define

$$D_t(\mathbf{c}) = \rho(y_t - \log(h^*_{t,k}(\mathbf{c}))) - g(u_{t-1}(\mathbf{c})) = \rho(w_t + u_{t-1}(\mathbf{c})) - g(u_{t-1}(\mathbf{c})),$$

where

$$u_{t-1}(\mathbf{c}) = \log(\widetilde{h}_t(\gamma)) - \log(h^*_{t,k}(\mathbf{c})).$$

It is easy to verify that $D_t(\mathbf{c})$ is a bounded martingale difference sequence. Then, by the law of large numbers for martingale differences, see for instance Theorem 20.10 of Davidson (1994), we get that

$$\lim_{T\to\infty} \frac{1}{T-p} \sum_{t=p+1}^{T} D_t(\mathbf{c}) = 0 \quad \text{a.s.} \tag{62}$$

Using a compactness argument for all $\varepsilon > 0$ we can find $(\boldsymbol{c}_i, \delta_i)$, $1 \leqslant i \leqslant K_0$, with $c_i \in C$, such that if we define

$$V_i = \{\mathbf{c} \in C : \|\mathbf{c} - \mathbf{c}_i\| \leqslant \delta_i\},$$

we have that $\cup_{i=1}^{K_0} V_i \supset C$ and

$$\sup_{\mathbf{c}\in V_i} \left| \frac{1}{T-p} \sum_{t=p+1}^{T} D_t(\mathbf{c}) - \frac{1}{T-p} \sum_{t=p+1}^{T} D_t(\mathbf{c}_i) \right| \leqslant \varepsilon.$$

This last inequality and (62) imply that

$$\lim_{T\to\infty} \sup \sup_{\mathbf{c}\in C} \left| \frac{1}{T-p} \sum_{t=p+1}^{T} D_t(\mathbf{c}) \right| \leqslant \varepsilon \quad \text{a.s.},$$

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

$$\lim_{T\to\infty} \sup \sup_{\mathbf{c}\in C} \left| \frac{1}{T-p} \sum_{t=p+1}^{T} D_t(\mathbf{c}) \right| = 0 \quad \text{a.s.} \tag{63}$$

Put

$$a_T = \frac{1}{T-p} \sum_{t=p+1}^{T} I_{A_t}.$$

Then, we get

$$\sup_{\mathbf{c}\in C} \frac{1}{T-p} \sum_{t=p+1}^{T} g(u_{t-1}(\mathbf{c})) = \sup_{\mathbf{c}\in C} \frac{1}{T-p} \sup_{t=p+1}^{T} g(u_{t-1}(\mathbf{c})) \geqslant (1-a_T)g(0) + a_T \inf_{u\geqslant k_1} g(u),$$

and therefore since $a_T \to a$ a.s., by (61) we have

$$\lim_{T\to\infty} \inf \sup_{\mathbf{c}\in C} \frac{1}{T-p} \sum_{t=p+1}^{T} g(u_{t-1}(\mathbf{c})) \geqslant (1-a)g(0) + a \inf_{u\geqslant k_1} g(u) \geqslant m_0 + \delta \quad \text{a.s.}$$

and then from (63) we have

$$\lim_{T\to\infty} \inf \sup_{\mathbf{c}\in C} \frac{1}{T-p} \sum_{t=p+1}^{T} \rho(y_t - \log(h^*_{t,k}(\mathbf{c}))) \geqslant m_0 + \delta.$$

Then the lemma follows. □

Proof of Theorem 4. Let $\widehat{\gamma}_2$ be as defined in (16). From Lemma 11 we have that $\liminf_{T\to\infty} M^*_{Tk}(\widehat{\gamma}_{2,T}) > m_0 + \delta$ a.s. for some $\delta > 0$. On the other hand, by Theorem 2 we have that $\widehat{\gamma}_T$ as defined in (13) satisfy $\lim_{T\to\infty} M_T(\widehat{\gamma}_T) = m_0$. This proves the theorem. □

References

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

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

Bianco, A.M., García Ben, M., Martínez, E.J., Yohai, V.J., 1996. Robust procedures for regression models with ARIMA errors. In: Prat, A. (Ed.), COMPSTAT 96, Proceedings in Computational Statistics. Physica-Verlag, Heidelberg, pp. 27–38.

Bianco, A.M., García Ben, M., Yohai, V.J., 2005. Robust estimation for linear regression with asymmetric errors. Canad. J. Statist. 33, 511–528.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.

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

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

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

Christian, F., Zakoïan, J.M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10, 605–637.

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

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

Franses, P.H., van Dijk, D., 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 Rev. 5, 57–61.

Hall, P., Yao, Q., 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. Ann. Math. Statist. 35, 73–101.

Huber, P.J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Statist. 1, 799–821.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

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

Li, J., Kao, C., 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.

Martin, R.D., Samarov, A., Vandaele, W., 1983. Robust methods for ARIMA models. In: Zellner, E. (Ed.), Applied Time Series. Analysis of Economic Data. Bureau of the Census, Washington, DC, pp. 153–177.

Mendes, B.V.M., 2000. Assessing the bias of maximum likelihood estimates of contaminated GARCH Models. J. Statist. Comput. Simulation 67, 359–376.

Mendes, B.V.M., Duarte, A.M., 1999. Robust estimation for ARCH models. Rev. Econometria 19, 138–180.

Muler, N., Yohai, V.J., 2002. Robust estimates for ARCH Processes. J. Time Series Anal. 23, 341–375.

Muler, N., Yohai, V.J., 2007. Robust estimates for GARCH models. Technical Report, Instituto de Calculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Available at: ⟨http://mate.dm.uba.ar/∼vyohai/muler-yohai-GARCH⟩.

Nelson, D., 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. J. Forecasting 21, 381–393.

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

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

Ronchetti, E., Trojani, F., 2001. Robust inference with GMM estimators. J. Econometrics 101, 37–69.

Sakata, S., White, H., 1998. High breakdown point conditional dispersion estimation with application to S&P daily returns volatility. Econometrica 66, 529–567.

Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equation approach. Ann. Statist. 34, 2449–2495.

van Dijk, D., Lucas, A., Franses, P.H., 1999. Testing for ARCH in the presence of additive outliers. J. Appl. Econometrics 14, 539–562.

Yohai, V.J., 1987. High breakdown-point and high efficiency robust estimates for regression. Ann. Statist. 15, 642–656.