doi:10.1111/j.1467-9892.2006.00519.x

EFFECTS OF OUTLIERS ON THE IDENTIFICATION AND ESTIMATION OF GARCH MODELS

By M. Angeles Carnero, Daniel Peña and Esther Ruiz

Universidad de Alicante and Universidad Carlos III de Madrid

First Version received July 2005

Abstract. This paper analyses how outliers affect the identification of conditional heteroscedasticity and the estimation of generalized autoregressive conditionally heteroscedastic (GARCH) models. First, we derive the asymptotic biases of the sample autocorrelations of squared observations generated by stationary processes and show that the properties of some conditional homoscedasticity tests can be distorted. Second, we obtain the asymptotic and finite sample biases of the ordinary least squares (OLS) estimator of ARCH(p) models. The finite sample results are extended to generalized least squares (GLS), maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators of ARCH(p) and GARCH(1,1) models. Finally, we show that the estimated asymptotic standard deviations are biased estimates of the sample standard deviations.

Keywords. Autocorrelations; generalized least squares heteroscedasticity; maximum likelihood; McLeod-Li test; ordinary least squares.

1. INTRODUCTION

Autoregressive conditional heteroskedastic (ARCH) models were introduced by Engle (1982) and extended to generalized ARCH (GARCH) by Bollerslev (1986) to represent the dynamic evolution of conditional variances. However, when these models are fitted to real time series, the residuals often have excess kurtosis, which could be explained, among other reasons, by the presence of outliers.

As in linear models, outliers affect the identification and estimation of GARCH models. It is known that outliers may wrongly suggest conditional heteroscedasticity. Van Dijk *et al.* (1999) analyse the properties of the Lagrange multiplier (LM) test for ARCH models in the presence of isolated additive outliers and show that, when the conditional mean has an autoregressive component, the LM test rejects the true null hypothesis of conditional homoscedasticity too often. Similar conclusions have been obtained by other authors analysing real time series of macroeconomic and financial variables (see, e.g. Balke and Fomby, 1994; Franses and Ghijsels, 1999; Aggarwal *et al.*, 1999; Franses *et al.*, 2004). It is also known that outliers may hide true heteroscedasticity (see Van Dijk *et al.*, 1999; Mendes, 2000; Li and Kao, 2002).

With respect to estimation, Sakata and White (1998) and Mendes (2000) analyse the finite sample effects of neglecting a single outlier on the maximum likelihood (ML) estimator of the parameters of GARCH(1,1) models. The conclusion in both papers is that the parameter measuring the ARCH effect is biased towards zero while the biases of the parameter related to the persistence of volatility are not clear and depend on the sample size. Furthermore, they also observe a loss of precision in the estimation of all the parameters.

This article extends these analyses in several directions. First of all, we derive the asymptotic biases caused by outliers on the sample autocorrelations of squared observations generated by stationary processes. We show that additive outliers in uncorrelated stationary series bias the sample autocorrelations of squares in the same direction, regardless of whether the generating process is homoscedastic or heteroscedastic. When outliers appear in patches, the autocorrelations of squares are different from zero, while when they are isolated, the autocorrelations are zero. The biases of the sample autocorrelations are then used to analyse the effects of outliers on the size and power of some popular homoscedasticity tests. We also carried out extensive Monte Carlo experiments to study which sizes of the outliers are expected to have significant effects on the testing results. In particular, we show that, if the sample size is large enough, relatively small consecutive outliers can lead the tests to detect spurious conditional heteroscedasticity, while isolated outliers hide genuine heteroscedasticity only if they are very large. We also analyse a robust test for conditional heteroscedasticity proposed by Van Dijk *et al.* (1999) and show that, in large samples, its size is distorted.

With respect to estimation, we obtain the asymptotic biases of the ordinary least squares (OLS) estimator of the parameters of ARCH(p) models and analyse their finite sample behaviour by means of extensive Monte Carlo experiments. Interestingly, we show that in finite samples, outliers can generate negative estimates of the ARCH parameters violating the restrictions for the positivity of the conditional variance. The finite sample results are extended to generalized least squares (GLS) and ML estimators of ARCH(p) and GARCH(1,1) models. We show that the GLS estimator is more robust than the OLS and is similar to ML. We also analyse the finite sample behaviour, in the presence of outliers, of a quasi-maximum likelihood (QML) estimator based on maximizing the Student t likelihood when the conditional distribution is truly Gaussian. We show that this estimator is resistant against outliers even when the sample is moderate and the outliers are relatively large. The properties of a closed-form estimator of the parameters of GARCH(1,1) models recently proposed by Kristensen and Linton (2006) have also been analysed. Given that this estimator is based on the sample autocorrelations of squared observations, the biases of these autocorrelations caused by outliers affect the properties of the closed-form estimator. Finally, we also study the effects of outliers on the estimated asymptotic standard deviations of the estimators considered and show that they are sometimes biased estimates of the sample standard deviations. Consequently, the inference on the GARCH parameters can be seriously affected, even in the presence of outliers of moderate sizes.

The article is organized as follows. Section 2 derives the biases caused by additive outliers on the sample autocorrelations of squared observations. Section 3 analyses how these biases affect the size and power of conditional homoscedasticity tests. Section 4 deals with the asymptotic and finite sample biases of the OLS, GLS, ML and QML estimators of the parameters of ARCH(p) models contaminated by additive outliers. These results are also extended to the ML and QML estimators of the parameters of GARCH(1,1) models in Section 5. Finally, Section 6 concludes the paper.

2. EFFECTS OF OUTLIERS ON THE SAMPLE AUTOCORRELATIONS OF SQUARED OBSERVATIONS

In this section, we derive analytically the effect of large additive outliers on the sample autocorrelations of squared observations generated by stationary processes that could be either homoscedastic or heteroscedastic.

Consider that the series of interest, y_t, $t = 1, \ldots, T$, is a stationary series with finite fourth-order moment that is contaminated from time τ onwards by k consecutive outliers of the same size ω. The observed series is given by

$$z_t = \begin{cases} y_t + \omega & \text{if } t = \tau, \tau+1, \ldots, \tau+k-1 \\ y_t & \text{otherwise.} \end{cases} \tag{1}$$

In this article, we focus on additive outliers because our interest is in the analysis of daily financial returns which are often characterized by being uncorrelated. In this context, the traditional distinction between additive and innovative outliers is not relevant. In any case, it is well known that the effects of innovative outliers on the dynamic properties of the series are less important as they are transmitted by the same dynamics as in the rest of the series (see, e.g. Peña, 2001). On the other hand, it is important to distinguish whether an outlier affects or not future conditional variances. We assume that the additive outliers defined in (1) are level outliers (LO) in the sense that they affect the level of the series but not the evolution of the underlying volatility [see Hotta and Tsay, 1998; Sakata and White, 1998 for the distinction between LO and volatility outliers (VO) in the context of GARCH models]. VO are defined in such a way that the underlying conditional variance depends on the observed series. Once more, we expect that similar to what happens in the context of linear models, the effects of VO are less important than those of LO.

The autocorrelation of order h, $h \geq 1$, of the squared observations of the contaminated series in eqn (1) is estimated by

$$r(h) = \frac{\sum_{t=h+1}^{T} z_t^2 z_{t-h}^2 - \frac{T-h}{T^2}\left(\sum_{t=1}^{T} z_t^2\right)^2}{\sum_{t=1}^{T} z_t^4 - T^{-1}\left(\sum_{t=1}^{T} z_t^2\right)^2}. \tag{2}$$

If the sample size, T, is large relative to the order of the estimated autocorrelation, h, the numerator of $r(h)$ can be written as follows

$$\sum_{t\in\mathbb{T}(h)} y_t^2 y_{t-h}^2 + \sum_{i=0}^{h-1}(y_{\tau+i}+\omega)^2 y_{\tau+i-h}^2 + \sum_{i=h}^{k-1}(y_{\tau+i}+\omega)^2(y_{\tau+i-h}+\omega)^2$$
$$+ \sum_{i=k}^{k+h-1} y_{\tau+i}^2(y_{\tau+i-h}+\omega)^2 - T^{-1}\left[\sum_{t\in\mathbb{T}(0)} y_t^2 + \sum_{i=0}^{k-1}(y_{\tau+i}+\omega)^2\right]^2 \tag{3}$$

where $\mathbb{T}(s) = \{s+1,\ldots,\tau-1,\tau+k+s,\ldots,T\}$. Similarly, the denominator can be written as

$$\sum_{t\in\mathbb{T}(0)} y_t^4 + \sum_{i=0}^{k-1}(y_{\tau+i}+\omega)^4 - T^{-1}\left[\sum_{t\in\mathbb{T}(0)} y_t^2 + \sum_{i=0}^{k-1}(y_{\tau+i}+\omega)^2\right]^2 \tag{4}$$

If the order of the autocorrelation is smaller than the number of consecutive outliers, i.e. $h < k$, then the third summation in (3) contains $k - h$ terms which depend on ω^4. Therefore, eqn (3) is equal to $(k - h - (k^2/T))\omega^4 + o(\omega^4)$. However, if $h \geq k$ then the third summation in (3) disappears and the numerator of $r(h)$ is equal to $-(k^2/T)\omega^4 + o(\omega^4)$. On the other hand, equation (4) is equal to $(k - (k^2/T))\omega^4 + o(\omega^4)$. Then

$$\lim_{\omega\to\infty} r(h) = \begin{cases} 1 - \frac{h}{k(1-\frac{k}{T})} & \text{if } h < k \\ \frac{k}{k-T} & \text{if } h \geq k \end{cases} \tag{5}$$

Therefore, one single large outlier ($k = 1$) always biases towards zero $r(h)$ for all lags, while a set of k large consecutive outliers generates positive $r(h)$ for $h < k$ and zero for the others. Furthermore, for $h < k$ and large T, the autocorrelations follow a linear decay for large outliers. For example, two large consecutive outliers generate an autocorrelation of the squares of order one approximately equal to 0.5, all the others being close to zero. Thus, if a heteroscedastic series is contaminated by a large single outlier, the detection of genuine heteroscedasticity will be difficult. On the other hand, when a homoscedastic series is contaminated by several large consecutive outliers, the positive autocorrelations of squares generated by the outliers can be confused with conditional heteroscedasticity.

It is important to note that the limits in (5) are valid regardless of whether y_t is a homoscedastic or heteroscedastic process. Moreover, note that although we have assumed that the outliers have the same sign, the limiting result only depends on ω^4 and would be the same if the signs are different. In addition, we can allow for different sizes and write ω_t instead of ω in eqn (1) and the results will be the same as far as all the ω_t go equally fast to infinity.

3. EFFECTS OF OUTLIERS ON TESTING FOR CONDITIONAL HETEROSCEDASTICITY

Many popular tests for conditional homoscedasticity, such as those proposed by Engle (1982), McLeod and Li (1983), Peña and Rodriguez (2002), and Rodriguez and Ruiz (2005) among others, are based on autocorrelations of squares and if these autocorrelations are biased, their properties will be affected. In this section, we analyse the behavior of such tests in the presence of outliers. As an example, we will focus on the McLeod and Li test which uses the Box–Ljung statistic for squared observations given by

$$Q(m) = T(T+2)\sum_{j=1}^{m}\frac{r^2(j)}{(T-j)}.$$

Under the null hypothesis of conditional homoscedasticity, if the eighth-order moment of y_t exists, $Q(m)$ is approximately distributed as a chi-squared distribution with m degrees of freedom.

On the other hand, Engle (1982) proposed a Lagrange multiplier (LM) test of homoscedasticity which is asymptotically equivalent to the McLeod and Li test.[1] Van Dijk *et al.* (1999) investigate the properties of the LM test in the presence of isolated additive outliers. In particular, they show that, in the presence of large isolated outliers, the size of the LM test when implemented to the residuals of an AR(1) model is larger than the nominal. On the other hand, in this case, there is an asymptotic power loss of the LM test when implemented to GARCH white-noise series (see also, Lee and King, 1993). They propose an alternative robust version of the LM test (RLM) which has better size and power properties in the presence of outliers. The RLM(m) statistic is given by TR^2 where R^2 is the determination coefficient when regressing $\psi(r_t)^2$ on a constant and $\psi(r_{t-1})^2, \ldots, \psi(r_{t-m})^2$ where

$$\psi(x) = x(1 - H(|x| - c_1))\mathrm{sign}(x) + H(|x| - c_1)\big(1 - H(|x| - c_2)\big)g(|x|) \qquad (6)$$

with c_1 and c_2 being constants (the authors consider $c_1 = 2.576$, $c_2 = 3.291$), $H(x) = I(x > 0)$, $sign(\cdot)$ is the sign function and $g(x)$ is an order 5 polynomial that makes the ψ function twice differentiable. On the other hand,

$$r_t = \frac{y_t}{\sigma_y \omega_y(y_{t-1})}$$

with σ_y being the MAD of y_t and

$$\omega_y(y_{t-1}) = \frac{\psi(d(y_{t-1})^2)}{d(y_{t-1})^2}$$

where $d(\cdot)$ is given by

$$d(y_{t-1}) = \frac{|y_{t-1} - m_y|}{\sigma_y}$$

and m_y is the median of y_t. The RLM(m) statistic is asymptotically chi-squared-distributed with m degrees of freedom.

We first analyse the properties of the McLeod–Li test when the series y_t is affected by a isolated large outlier. In this case, from eqn (5), the limit of the estimated autocorrelations of any order is zero, so that the null is never rejected. Thus, if the series is homoscedastic the size is zero while if the series is heteroscedastic, the power is also zero.

When the series is affected by k consecutive outliers, from eqn (5) we know that the limit of the order-one autocorrelation is $1 - (T/k(T-k))$. Then,

$$\lim_{\omega\to\infty} Q(1) = \frac{T(T+2)}{(T-1)}\left(1 - \frac{T}{k(T-k)}\right)^2 \underset{T\to\infty}{\longrightarrow} \infty,$$

and the null will always be rejected. Thus, if the series is truly homoscedastic, the asymptotic size is, in this case, one. On the other hand, if the series is heteroscedastic, the power is also one. The same kind of arguments can be used to show that the asymptotic size and power of the LM, Peña–Rodriguez and Rodriguez–Ruiz tests are zero when series are contaminated by a large isolated outlier, while they are one in the presence of large consecutive outliers.

To analyse the finite sample effects of moderate outliers on these tests, we simulated 1000 Gaussian white-noise series of sizes $T = 500$, 1000 and 5000 that have been contaminated, at time $\tau = T/2$, first by one single outlier and then, by two consecutive outliers of the same size ω. For each simulated series, we test the null hypothesis of conditional homoscedasticity using the $Q(20)$ and the RLM(20) tests. The top panel on the left of Figure 1 plots the empirical sizes of both tests as a function of the outlier size when it is isolated and the nominal size is 5%. This plot shows that, for $T = 500$ or 1000, the size of $Q(20)$ is zero for outliers larger than 8 standard deviations while the size of RLM(20) is around the nominal, 5%, regardless of the outlier size. However, when $T = 5000$, the size of $Q(20)$ tends to zero, only if the outlier is larger than approximately 12 standard deviations while the size of RLM(20) is around 9%, i.e. nearly double the nominal, independently of the outlier size. The robust test is oversized in large samples. Lee and King (1993) find similar size distortions in the robust test proposed by Wooldridge (1990).

The right panel on top of Figure 1 plots the empirical sizes of both tests when the Gaussian series are contaminated by two consecutive outliers. In this case, the behaviour of the robust test is similar to the one observed when there is just one outlier. However, for relatively small outlier sizes, like for example, 5 marginal standard deviations, the size of the non-robust tests is almost 1 for any of the three sample sizes considered. Therefore, rather small consecutive outliers in homoscedastic series make the McLeod–Li test detect conditional heteroscedasticity even for relatively large samples.

To analyse the power of both tests in the presence of additive LO, we generated series by the GARCH (1,1) model given by $y_t = \varepsilon_t\sigma_t$; $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta\sigma_{t-1}^2$;

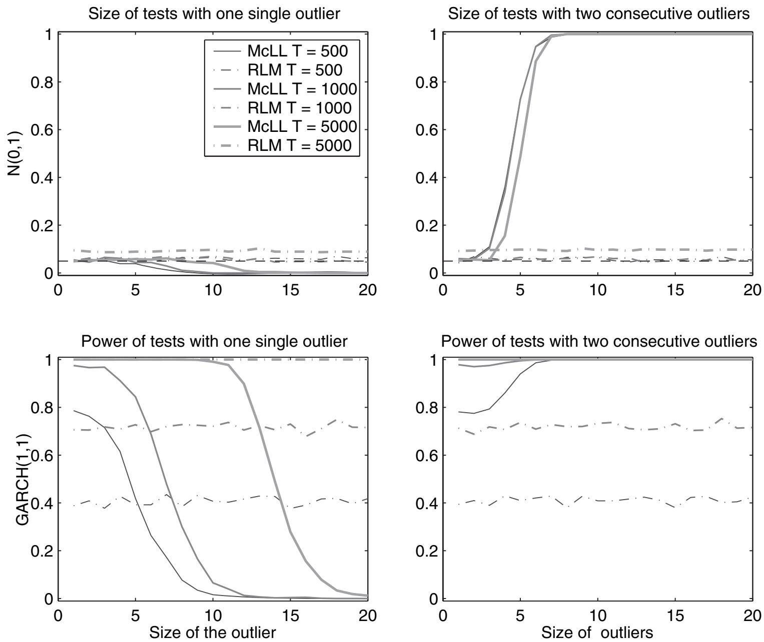

FIGURE 1. Effects caused by outliers on the size and power of conditional homoscedasticity tests.

where ε_t is a Gaussian white noise with mean zero and variance one and the parameters α_0, α_1 and β have been chosen to resemble the values usually estimated with real time series of financial returns. In particular, we chose $\alpha_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta = 0.8$ which satisfy the restrictions to guarantee the positiveness, stationary and existence of the fourth-order moment of y_t (see, e.g. Bollerslev *et al.*, 1994).

The power of the tests for isolated outliers is shown in the left bottom panel of Figure 1 as a function of the size of the outlier. This figure shows that if the outlier size is smaller than approximately 5 standard deviations, the power of the portmanteau test is larger than the power of the robust test when the sample size is $T = 500$ or 1000, respectively. For these sample sizes, the power of the $Q(20)$ test decreases rapidly with the size of the outlier. If this size is larger than approximately 10 standard deviations, the power is negligible. However, if $T = 5000$, a very large outlier is needed for the RLM(20) test to have more power than the $Q(20)$ test. In our experiments, the power of the $Q(20)$ test is affected only if the outlier is larger than 13 standard deviations. We have also contaminated the GARCH series with two consecutive outliers. The empirical powers have been plotted in the right bottom panel of Figure 1. For all sample sizes and outlier sizes chosen, the power of the robust test is clearly lower than that of the non-robust test considered. A similar result is obtained by Lee and King (1993) comparing the power of the robust test proposed by Wooldridge (1990) with the LM test.

Summarizing, when standard tests are used in testing for conditional homoscedasticity, relatively small consecutive outliers are able to generate spurious heteroscedasticity, while large isolated outliers are required to hide genuine heteroscedasticity. On the other hand, the available robust LM test could suffer from important size distortions especially in case of large sample sizes.

4. EFFECTS OF OUTLIERS ON THE ESTIMATION OF ARCH MODELS

The ARCH(p) model often requires a large number of lags, p, to adequately represent the dynamic evolution of the conditional variances. However, this model is attractive because it is possible to obtain a closed-form expression for the OLS estimator of its parameters. In the following Section 4.1, we quantify the effects of level outliers on the OLS estimator of ARCH(p) models. In Section 4.2, we also analyse the effects of outliers on the GLS estimator. Finally, the results are extended in the next subsections to ML and QML estimators.

4.1. *OLS estimator*

The ARCH(p) model is given by

$$y_t = \varepsilon_t \sigma_t \quad \text{where } \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2,$$

ε_t is a Gaussian white noise and the parameters α_i should be restricted so that σ_t^2 is positive and y_t is stationary with finite fourth-order moment. The ARCH(p) model is an AR(p) for squared observations given by

$$y_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + v_t$$

where the noise, $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$, is a zero-mean uncorrelated sequence. However, it is conditionally heteroscedastic and, consequently, is non-independent and non-Gaussian.

The OLS estimator of the parameters of the ARCH(p) model is given by

$$\hat{\boldsymbol{\alpha}}^{\text{OLS}} = (X'X)^{-1}(X'\mathbf{Y}_{p+1})$$

where $\boldsymbol{\alpha} = (\alpha_0 \quad \alpha_1 \ldots \alpha_p)'$, $\quad \mathbf{Y}_{p+1} = (y_{p+1}^2 \quad y_{p+2}^2 \ldots y_T^2)'$ and $X = (\mathbf{1} \quad \mathbf{Y}_p \cdots \mathbf{Y}_1)$

where $\mathbf{1}$ is a column vector of ones. Weiss (1986) shows that if the fourth-order moment of y_t exists, $\hat{\boldsymbol{\alpha}}^{\text{OLS}}$ is consistent (see Engle, 1982, for sufficient conditions for the existence of higher moments of y_t when ε_t is Gaussian). Furthermore, if the eighth-order moment is finite, the asymptotic distribution of $\hat{\boldsymbol{\alpha}}^{\mathrm{OLS}}$ is given by

$$\sqrt{T}(\hat{\boldsymbol{\alpha}}^{\mathrm{OLS}} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, \Sigma_{XX}^{-1}\Sigma_{X\Omega X}\Sigma_{XX}^{-1})$$

where

$$P\lim \frac{X'X}{T} = \Sigma_{XX} \quad \text{and} \quad P\lim \frac{X'VV'X}{T} = \Sigma_{X\Omega X} \;\; \text{with} \quad V = (v_{p+1}^2 v_{p+2}^2 \dots v_T^2)'$$

and $P\lim(x) = c$ meaning that x converges in probability to c.

A consistent estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}^{\mathrm{OLS}}$ is given by

$$(X'X)^{-1}S(X'X)^{-1} \tag{7}$$

where

$$S = \begin{pmatrix} \sum_{t=p+1}^{T} \hat{v}_t^2 & \sum_{t=p+1}^{T} \hat{v}_t^2 y_{t-1}^2 & \cdots & \sum_{t=p+1}^{T} \hat{v}_t^2 y_{t-p}^2 \\ & \sum_{t=p+1}^{T} \hat{v}_t^2 y_{t-1}^4 & \cdots & \sum_{t=p+1}^{T} \hat{v}_t^2 y_{t-1}^2 y_{t-p}^2 \\ & & \ddots & \vdots \\ & & & \sum_{t=p+1}^{T} \hat{v}_t^2 y_{t-p}^4 \end{pmatrix}$$

and $\hat{v}_t$ are the residuals from the OLS regression. Next, we analyse how a single outlier affects the asymptotic properties of $\hat{\boldsymbol{\alpha}}^{\mathrm{OLS}}$. We then consider the effects of patches of outliers.

4.1.1. *Isolated outliers*

Consider a series generated by an ARCH(p) model which is contaminated at time τ by a single level outlier of size ω, as in eqn (1) with $k = 1$. Then, $\hat{\boldsymbol{\alpha}}^{\mathrm{OLS}}$ will be computed using the contaminated observations z_t^2 instead of y_t^2 by

$$\begin{pmatrix} \hat{\alpha}_0^{\mathrm{OLS}} \\ \hat{\alpha}_1^{\mathrm{OLS}} \\ \vdots \\ \hat{\alpha}_p^{\mathrm{OLS}} \end{pmatrix} = \begin{pmatrix} T-p & \sum_{t=p}^{T-1} z_t^2 & \cdots & \sum_{t=p}^{T-1} z_{t-p+1}^2 \\ \sum_{t=p}^{T-1} z_t^2 & \sum_{t=p}^{T-1} z_t^4 & \cdots & \sum_{t=p}^{T-1} z_t^2 z_{t-p+1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=p}^{T-1} z_{t-p+1}^2 & \sum_{t=p}^{T-1} z_t^2 z_{t-p+1}^2 & \cdots & \sum_{t=p}^{T-1} z_{t-p+1}^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=p+1}^{T} z_t^2 \\ \sum_{t=p+1}^{T} z_t^2 z_{t-1}^2 \\ \vdots \\ \sum_{t=p+1}^{T} z_t^2 z_{t-p}^2 \end{pmatrix} \tag{8}$$

Taking into account that $z_\tau^2 = \omega^2 + o(\omega^2)$, $z_\tau^4 = \omega^4 + o(\omega^4)$ and $z_t^r = o(\omega)$ for $t \neq \tau$ and $\forall r \geq 0$, the matrix $X'X$ can be written as

$$\begin{pmatrix} T-p & (\omega^2 + o(\omega^2))\mathbf{1}' \\ (\omega^2 + o(\omega^2))\mathbf{1} & (\omega^2 + o(\omega^2))\mathbf{F} \end{pmatrix}$$

and $(X'X)^{-1}$ can be written as

$$\frac{1}{(T-2p)\omega^{4p}}\begin{pmatrix} \omega^{4p}+o(\omega^{4p}) & (-\omega^{4p-2}+o(\omega^{4p-2}))\mathbf{1}' \\ (-\omega^{4p-2}+o(\omega^{4p-2}))\mathbf{1} & \mathbf{V} \end{pmatrix}$$

where $\mathbf{F}$ is a $p \times p$ symmetric matrix with $f_{ii} = \omega^2$ for $i = 1,\ldots,p$ and all other elements are equal to one. $\mathbf{V}$ is a $p \times p$ symmetric matrix with all its elements equal to $o(\omega^{4p-2})$. Finally, all elements in $X'\mathbf{Y}_{p+1}$ are equal to $\omega^2 + o(\omega^2)$. Thus,

$$\lim_{\omega\to\infty} \hat{\boldsymbol{\alpha}}^{\mathrm{OLS}} = \lim_{\omega\to\infty} \frac{1}{(T-2p)\omega^{4p}}\begin{pmatrix} \omega^{4p+2}+o(\omega^{4p+2}) \\ (-\omega^{4p}+o(\omega^{4p}))\mathbf{1} \end{pmatrix} = \begin{pmatrix} \infty \\ -\frac{1}{T-2p}\mathbf{1} \end{pmatrix}. \qquad (9)$$

The limit in (9) shows that, if the sample size is large enough and the outlier size goes to infinity, all the estimated ARCH parameters tend to zero. Consequently, the dynamic dependence in the conditional variance disappears. Notice that the persistence of the volatility in an ARCH(p) model, measured by $\sum_{i=1}^{p}\alpha_i$, also decreases as the size of the outlier increases and obviously, the estimated unconditional variance, given by $\hat{\alpha}_0/(1-\sum_{i=1}^{p}\hat{\alpha}_i)$, tends to infinity. Finally, it is also important to notice that if the sample size is not very large, it is possible to obtain estimates that do not satisfy the usual non-negativity restrictions (see, e.g. the simulation results in Mendes, 2000).

4.1.2. *Patches of outliers*

When the original series, y_t, is contaminated by k consecutive outliers, the effects on the OLS estimator depend on the relationship between the number of outliers and the order of the ARCH model. Let us consider $k \geq p$, i.e. that there are at least as many outliers as the number of lags in the ARCH model. In this case, it is necessary to consider the cases separately where $p = 1$ and $p > 1$. This is because in the first case, the parameter α_1 receives the whole effect of the outliers while in the latter, this effect is shared by all the parameters.

We first consider the effect of k consecutive outliers of size ω on the OLS estimates of the parameters of an ARCH(1) model. In this case,

$$\sum_{t=1}^{T-1} z_t^2 = k\omega^2 + o(\omega^2) \quad \text{and} \sum_{t=1}^{T-1} z_t^4 = k\omega^4 + o(\omega^4)$$

and the following result is obtained

$$\lim_{\omega\to\infty} \hat{\boldsymbol{\alpha}}^{\mathrm{OLS}} = \lim_{\omega\to\infty} \frac{1}{((T-1)k-k^2)\omega^4}\begin{pmatrix} k\omega^4+o(\omega^4) & -k\omega^2+o(\omega^2) \\ -k\omega^2+o(\omega^2) & T-1 \end{pmatrix} \times \begin{pmatrix} k\omega^2+o(\omega^2) \\ (k-1)\omega^4+o(\omega^4) \end{pmatrix}.$$

Hence,

$$\lim_{\omega\to\infty} \hat{\alpha}_i^{\mathrm{OLS}} = \begin{cases} \infty & \text{for } i=0 \\ \frac{(T-1)(k-1)-k^2}{(T-1)k-k^2} & \text{for } i=1 \end{cases} \qquad (10)$$

Notice that if $k = 1$, we obtain the same result as in eqn (9). If the number of consecutive outliers is large, the estimated ARCH parameter, $\hat{\alpha}_1$, tends to one when ω tends to infinity. Therefore, the presence of long patches of large outliers can lead us, to infer that the conditional variance has a unit root and, consequently, that y_t is not stationary. Notice that patches of large outliers can overestimate or underestimate the ARCH parameter depending on its original value. For example, if the sample size is moderate and there are two large consecutive outliers, $\hat{\alpha}_1$ tends to 0.5. Therefore, if $\alpha_1 < 0.5$, the OLS estimator will have a positive bias while if $\alpha_1 > 0.5$, the bias will be negative. However, notice that in cases of empirical interest in the context of financial time series, the ARCH parameter is usually rather small, never over 0.3, and then with patches of consecutive outliers, the OLS estimator will overestimate the ARCH parameter. In particular, if the series is truly homoscedastic, i.e. $\alpha_1 = 0$, then the estimated ARCH parameter will be close to 0.5 and can lead us to conclude that the series is conditionally heteroscedastic. Finally, it is also important to point out that the limit in eqn (10) increases very quickly with the number of consecutive outliers. For example, if $k = 3$, $\hat{\alpha}_1$ tends to 0.66 while if $k = 4$ the limit is 0.75.

Next, we consider the effect of $k \geq p$ consecutive outliers in an ARCH(p) model with $p > 1$. Consider again the OLS estimator of the parameters of the ARCH(p) model. If the series is contaminated by k consecutive outliers, then

$$\sum_{t=p}^{T-1} z_t^2, \sum_{t=p}^{T-1} z_{t-1}^2, \ldots, \sum_{t=p}^{T-1} z_{t-p+1}^2 \text{ are equal to } k\omega^2 + o(\omega^2),$$

$$\sum_{t=p}^{T-1} z_t^4, \sum_{t=p}^{T-1} z_{t-1}^4, \ldots, \sum_{t=p}^{T-1} z_{t-p+1}^4 \text{are equal to } k\omega^4 + o(\omega^4)$$

and

$$\sum_{t=p}^{T-1} z_t^2 z_{t+1}^2 = (k-1)\omega^4 + o(\omega^4), \ldots, \sum_{t=p}^{T-1} z_t^2 z_{t-p+1}^2 = (k-p+1)\omega^4 + o(\omega^4).$$

Therefore, the $X'X$ matrix can be written as

$$\begin{pmatrix} T-p & (k\omega^2 + o(\omega^2))\mathbf{1}' \\ (k\omega^2 + o(\omega^2))\mathbf{1} & (\omega^4 + o(\omega^4))\mathbf{M} \end{pmatrix}$$

where $\mathbf{M}$ is a $p \times p$ symmetric matrix with $m_{ij} = k + i - j$ for $i = 1, \ldots, p, j = i, \ldots, p$. Consequently, the OLS estimator is given by

$$\begin{pmatrix} \frac{2^{p-1}(k-(p-1)/2)}{2^{p-2}(-2k^2+(2k-p+1)(T-p))} & -\frac{1}{\omega^4}\frac{k}{-2k^2+(2k-p+1)(T-p)}\ell' \\ -\frac{1}{\omega^4}\frac{k}{-2k^2+(2k-p+1)(T-p)}\ell & (\omega^4 + o(\omega^4))\mathbf{D} \end{pmatrix} \begin{pmatrix} k\omega^2 + o(\omega^2) \\ (\omega^4 + o(\omega^4))\mathbf{B} \end{pmatrix}$$

where $\mathbf{D}$ is a $p \times p$ symmetric matrix with

$$d_{11} = d_{pp} = \frac{1}{2\omega^4}\frac{-2k^2 + (2k-p+2)(T-p)}{-2k^2 + (2k-p+1)(T-p)},$$

$$d_{ii} = \frac{1}{\omega^4} \quad \text{for } i = 2, \ldots, p-1,$$

$$d_{ii+1} = -\frac{1}{2\omega^4} \quad \text{for } i = 2, \ldots, p-1,$$

$$d_{1p} = \frac{1}{2\omega^4}\frac{T-p}{-2k^2 + (2k-p+1)(T-p)} \quad \text{and} \quad d_{ij} = 0 \text{ otherwise},$$

$\ell = (1\ 0\ 0 \cdots 0\ 1)'$ and $\mathbf{B}$ is a $p \times 1$ column vector such that $b_i = k - i$ for $i = 1, \ldots, p$. Then,

$$\lim_{\omega\to\infty} \hat{\alpha}_i^{\mathrm{OLS}} = \begin{cases} \infty & \text{for } i = 0 \\ \dfrac{-2k^2 + (2k-p)(T-p)}{-2k^2 + (2k-p+1)(T-p)} & \text{for } i = 1 \\ 0 & \text{for } i = 2, \ldots, p-1 \\ \dfrac{-(T-p)}{-2k^2 + (2k-p+1)(T-p)} & \text{for } i = p \end{cases} \tag{11}$$

The estimated parameters, $\hat{\alpha}_i$, tend to zero, except $\hat{\alpha}_1$ and $\hat{\alpha}_p$. If the number of consecutive outliers is large relative to the order of the model, then $\hat{\alpha}_1$ tends to a quantity close to one and $\hat{\alpha}_p$ tends to zero. Consequently, the estimated persistence, given by $\sum_{i=1}^{p} \hat{\alpha}_i$, tends to $(-2k^2 + (2k - p - 1)(T - p))/(-2k^2 + (2k - p + 1)(T - p))$ which is close to one. Notice that if $p = 1$, the limit of the persistence coincides with the limit of $\hat{\alpha}_1$ given in eqn (10). Consider, for example, an ARCH(2) series contaminated by two large consecutive outliers. In this case, if the sample size is moderately large, $\hat{\alpha}_1$ tends approximately to 0.66 and $\hat{\alpha}_2$ to -0.34 and, consequently, the persistence tends to 0.32. However, if the number of consecutive outliers is 5, $\hat{\alpha}_1$ tends to 0.89 and $\hat{\alpha}_2$ to -0.11 and the persistence tends to 0.78. On the other hand, if there are five consecutive outliers in an ARCH(4) series, $\hat{\alpha}_1$ tends to 0.86, $\hat{\alpha}_4$ to -0.15 and the persistence to 0.71. Again, in the presence of patches of outliers, the estimates may easily violate the non-negativity restrictions. Finally, the effect of $k < p$ consecutive outliers in an ARCH(p) model depends on the relationship between k and p (see Carnero, 2003, for particular cases).

4.2. *Generalized least squares estimator*

The OLS estimator is not efficient because the noise in the regression equation, v_t, is conditionally heteroscedastic. Bose and Mukherjee (2003) propose to estimate the parameters of ARCH(p) models using the GLS estimator which is computationally simple while having asymptotic efficiency equivalent to the ML estimator. The GLS estimator is obtained estimating by OLS the parameters $\boldsymbol{\alpha}$ in $PY = PX\alpha + PV$, where $P'P = \Omega^{-1}$ and $\Omega = \mathrm{diag}(\sigma_{p+1}^4, \ldots, \sigma_T^4)$. In practice given that the matrix Ω is unknown, it can be substituted by

$$\hat{\Omega} = \text{diag}(\hat{\sigma}^4_{p+1}, \ldots, \hat{\sigma}^4_T), \quad \text{where } \hat{\sigma}^2_t = \hat{\alpha}^{\text{OLS}}_0 + \hat{\alpha}^{\text{OLS}}_1 y^2_{t-1} + \cdots + \hat{\alpha}^{\text{OLS}}_p y^2_{t-p}.$$

Therefore, the GLS estimator is given by

$$\hat{\boldsymbol{\alpha}}^{\text{GLS}} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y),$$

and if the sixth-order moment of y_t is finite, the asymptotic distribution of $\hat{\boldsymbol{\alpha}}^{\text{GLS}}$ is given by

$$\sqrt{T}\left(\hat{\boldsymbol{\alpha}}^{\text{GLS}} - \boldsymbol{\alpha}\right) \xrightarrow{d} N(0, \Sigma_{\Omega}^{-1}) \tag{12}$$

where $P \lim(X'\Omega^{-1}X/T) = \Sigma_{\Omega}$.

Next, we compare the robustness of GLS and OLS estimators. Consider, for example, the case of the ARCH(1) model. When there is an isolated outlier both estimators are expected to be very similar because, as we have seen before, $\hat{\alpha}^{\text{OLS}}_1$ is biased towards zero and, consequently, the weights $\hat{\sigma}^{-4}_t$ for the GLS estimator will be almost constant. On the other hand, if there are consecutive outliers, $\hat{\alpha}^{\text{OLS}}_1$ is overestimated, the weights $\hat{\sigma}^{-4}_t$ down-weight the outliers and, therefore, the GLS estimator is expected to be more robust. To illustrate this result, we generate 1000 series of sizes $T = 500$, 1000 and 5000 by an ARCH(1) model with parameters $\alpha_0 = 0.8$ and $\alpha_1 = 0.2$. All the series have been contaminated with a single LO of size $\omega = 0,5,10$ and 15 marginal standard deviations. Figure 2 plots kernel estimates of the density of the OLS and GLS estimators of α_0 (rows 1 and 2) and α_1 (rows 5 and 6) all obtained through Monte Carlo replicates. Comparing the kernel densities of the estimators of α_0, we can observe that, as expected given the sample sizes considered in this article, both estimators are unbiased when there are no outliers. In this case, it is also possible to observe that the dispersion of the GLS estimator is smaller than the OLS estimator, especially for the largest sample size. On the other hand, in the presence of isolated outliers, both estimators have similar sample distributions with positive biases in small or moderate samples. However, when $T = 5000$, the bias of the GLS estimator is almost negligible even if $\omega = 15$, while the OLS estimator has large biases for rather small outliers. The performance of both estimators of α_1 is similar when there are no outliers. They are unbiased and GLS is more precise than OLS. However, in the presence of moderate isolated outliers, we can observe a large negative bias of the OLS estimator even if the sample size is large. On the other hand, when $T = 5000$, the GLS estimator of α_1 is unbiased in the presence of outliers as large as 15 standard deviations.

We also analyse how an isolated outlier affects the estimated variances of the OLS and GLS estimators. Figure 3 plots the logarithm of the ratio of the sample variance of $\hat{\alpha}^{\text{OLS}}_0$ and $\hat{\alpha}^{\text{OLS}}_1$ in the Monte Carlo experiments, and the estimated asymptotic variance computed as in eqn (7) and averaged through all Monte Carlo replicates. The corresponding quantities have also been computed for $\hat{\alpha}^{\text{OLS}}_0$ and $\hat{\alpha}^{\text{OLS}}_1$. As we can see in the graph (columns 1 and 2), the asymptotic variances of the OLS estimator overestimate the sample variances while the asymptotic variances of the GLS estimator underestimate them.

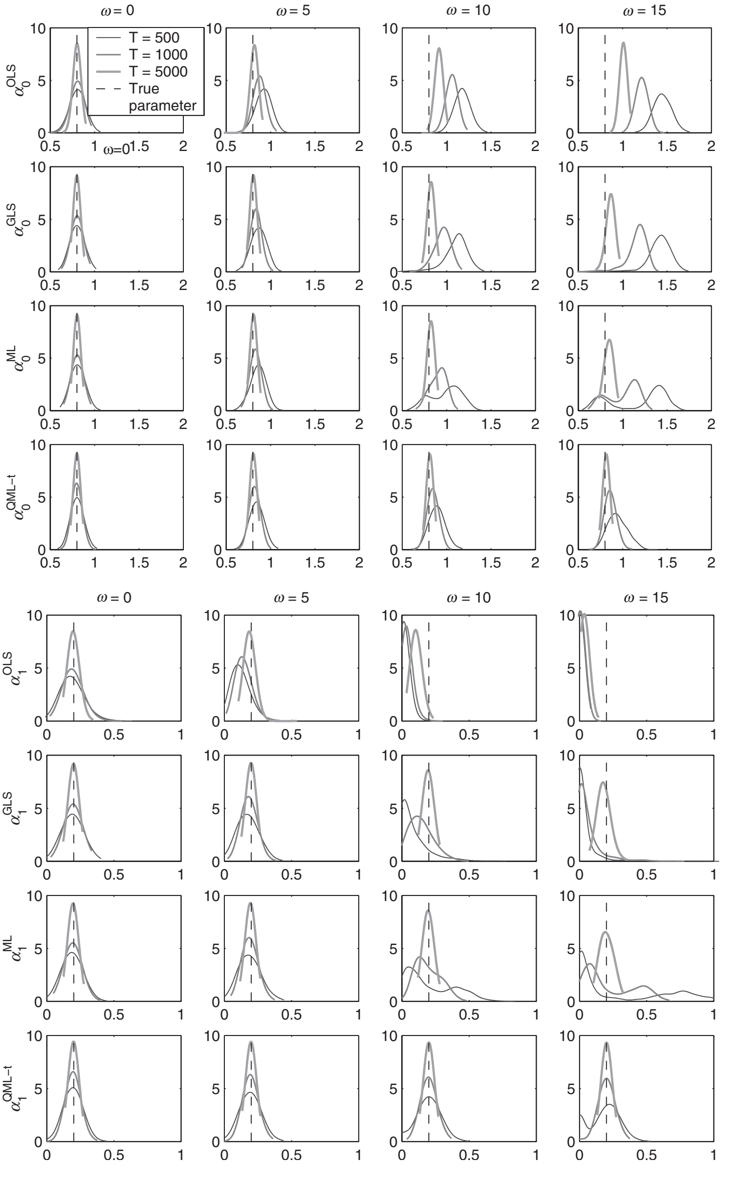

FIGURE 2. Kernel estimation of the density of OLS, GLS, ML and QML estimators of α_0 (rows 1 to 4) and α_1 (rows 5 to 8) in an ARCH(1) model with a single outlier of size ω.

To analyse the effect of consecutive outliers, each simulated series has also been contaminated by two consecutive outliers of the same sizes as above. Figure 4 plots kernel estimates of the densities of the OLS and GLS estimators of α_0 and α_1. It is important to note that although in the limit $\hat{\alpha}_0^{\text{OLS}}$ increases with ω, α_0 can be underestimated for small outliers. For example,

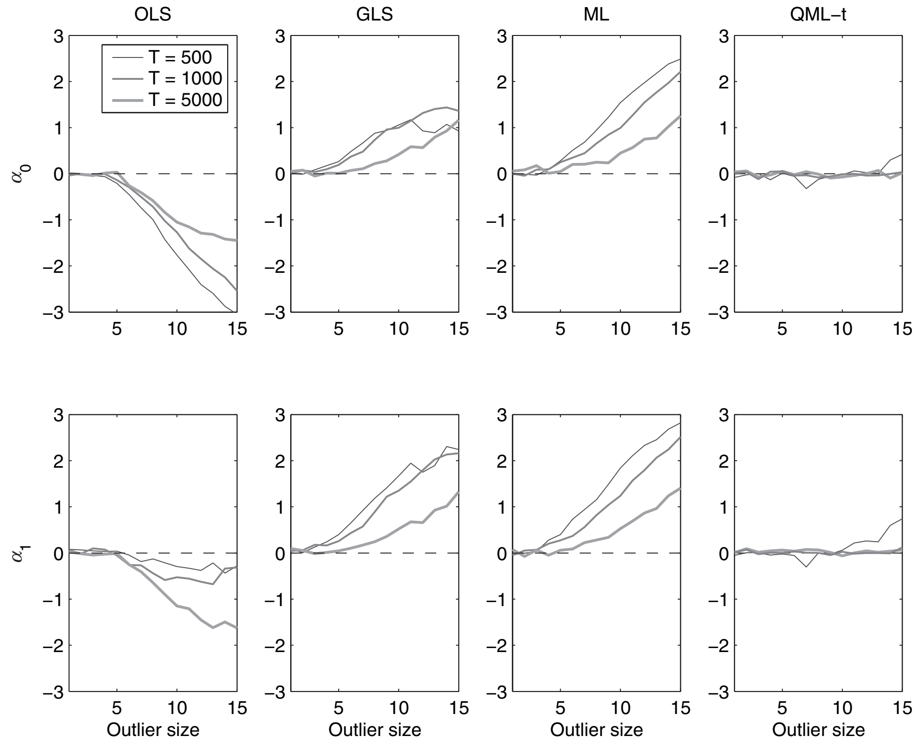

FIGURE 3. Logarithm of the ratio of variances of OLS, GLS, ML and QML estimators of α_0 (row 1) and α_1 (row 2) in an ARCH(1) model with a single outlier.

consider $T = 500$ or 1000, then if the outlier size is 5 marginal standard deviations, the mean of the estimates $\hat{\alpha}_0^{\mathrm{OLS}}$ is 0.75, below the true value of 0.8. However, if the size of the outlier is 15, the mean is 0.98. Consequently, for the outlier sizes typically encountered in empirical applications, the constant can be underestimated in the presence of patches of outliers. Remember that in the presence of a single outlier, the OLS estimates of α_0 tend monotonically to infinity. Therefore, although the effect in the limit is the same, in practice, isolated outliers overestimate the constant while consecutive outliers underestimate it. However, if the sample size is large enough, the bias of $\hat{\alpha}_0^{\mathrm{GLS}}$ is almost negligible for all the outlier sizes considered in this article. When the sample size is small or moderate, large consecutive outliers increase the dispersion of the $\hat{\alpha}_0^{\mathrm{GLS}}$ estimator in such a way that the inference is useless.

Looking at the results for $\hat{\alpha}_1^{\mathrm{OLS}}$, we observe that in concordance with the limit in eqn (10), they tend to 0.5 when $k = 2$. Furthermore, for all the sample sizes considered, the limit is reached for relatively small outliers. Once more, the bias of the GLS estimator of α_1 is almost negligible for $T = 5000$. However, if the sample size is moderate and the outliers are large, the dispersion of $\hat{\alpha}_1^{\mathrm{GLS}}$ is so large that inference is not reliable.

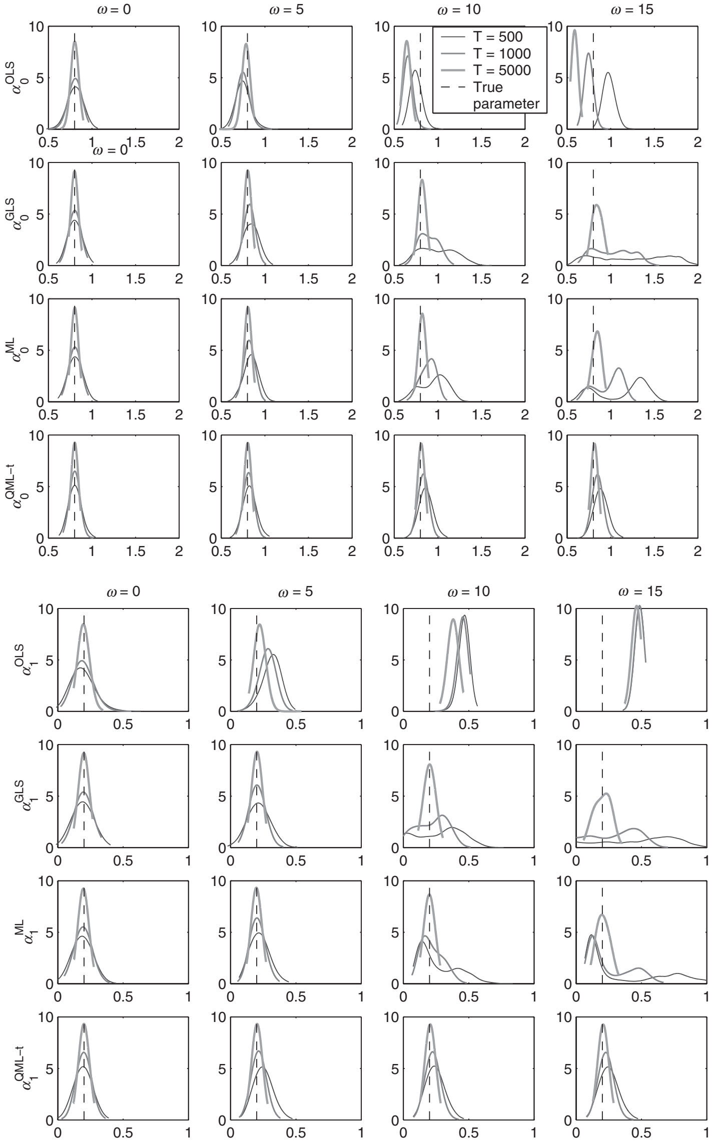

FIGURE 4. Kernel estimation of the density of OLS, GLS, ML and QML estimators of α_0 (rows 1 to 4) and α_1 (rows 5 to 8) in an ARCH(1) model with two consecutive outliers.

Finally, the logarithm of the ratio of the sample variance and the estimated asymptotic variance of the OLS and GLS estimators is plotted in Figure 5, where it is shown (see column 1) that, for the OLS estimator of α_0 and α_1, this ratio tends to zero with the size of the outlier. Therefore, the asymptotic variance of the OLS estimator, estimated using eqn (7), overestimates the true

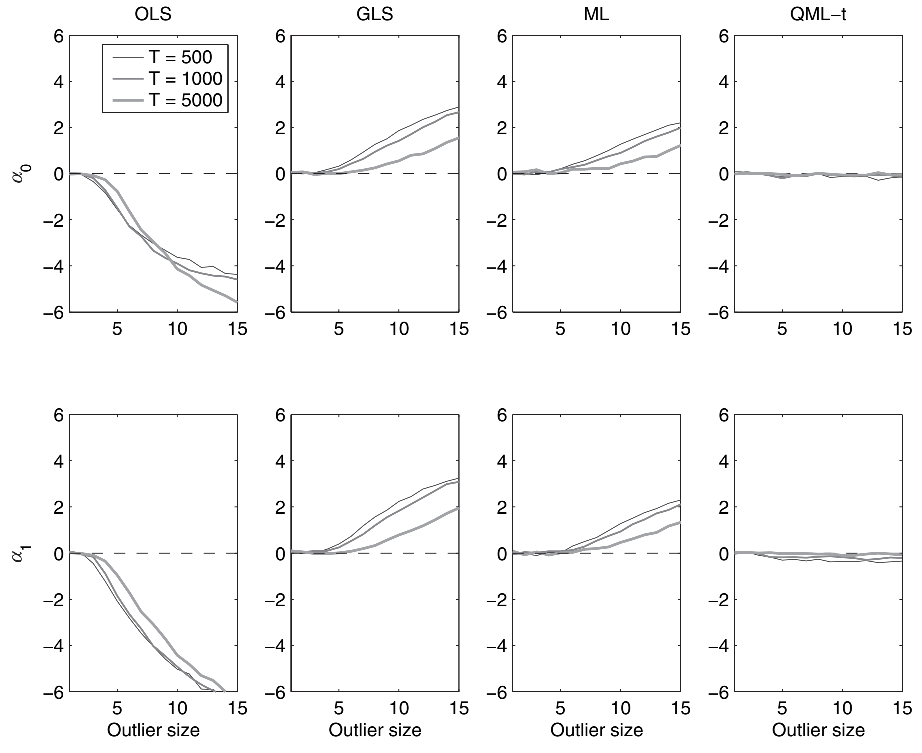

FIGURE 5. Logarithm of the ratio of variances of OLS, GLS, ML and QML estimators of α_0 (row 1) and α_1 (row 2) in an ARCH(1) model with two consecutive outliers.

variance, which tends to zero with the size of the outlier. Notice that in this case, the biases are larger than in the presence of a single outlier. moreover, the estimated asymptotic variances of the GLS estimator, strongly underestimate the sample variances for consecutive outliers larger than 5 standard deviations (see column 2).

4.3. *Maximum likelihood estimator*

Engle (1982) proposed to estimate the parameters of the ARCH(p) model by ML. The distribution of y_t conditional on $Y_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_1\}$ is $N(0, \sigma_t^2)$ and the log-likelihood function is given by

$$L = -\frac{T-p}{2}\log(2\pi) - \frac{1}{2}\sum_{t=p+1}^{T}\left(\log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2}\right). \tag{13}$$

If the second-order moment of y_t is finite, then,

$$\sqrt{T}(\hat{\boldsymbol{\alpha}}^{\mathrm{ML}} - \boldsymbol{\alpha}) \xrightarrow{d} N(0, [I(\alpha)]^{-1}) \quad \text{where } I(\alpha) = E[-\frac{\partial^2 L}{\partial \alpha \partial \alpha'}]$$

is the information matrix (see, e.g. Ling and McAleer, 2003).

Given that there are no closed-form expressions of the ML estimators of the parameters $\alpha_0, \alpha_1, \ldots, \alpha_p$, the analysis of the effects of outliers on the ML estimator has been carried out by simulation [see, e.g. Muller and Yohai, 2002, who show that the mean squared error of the ML estimator of the parameters of ARCH(1) models is dramatically influenced by isolated outliers]. On the other hand, Mendes (2000) shows that the influence functional of the ML estimator of the parameters of an ARCH(1) model is the product of a constant vector by a quadratic function of the outlier size. Consider the simplest ARCH(1) model. In this case, the log-likelihood function is given by

$$L = -\frac{T-2}{2}\log(2\pi) - \frac{1}{2}\sum_{t=2}^{T}\left(\log(\alpha_0 + \alpha_1 y_{t-1}^2) + \frac{y_t^2}{(\alpha_0 + \alpha_1 y_{t-1}^2)}\right)$$

which leads to the following ML equations to obtain the estimated parameters

$$\sum_{t=2}^{T}\frac{y_t^2}{\hat{\sigma}_t^4} = \sum_{t=2}^{T}\frac{1}{\hat{\sigma}_t^2}$$
$$\sum_{t=2}^{T}\frac{y_{t-1}^2 y_t^2}{\hat{\sigma}_t^4} = \sum_{t=2}^{T}\frac{y_{t-1}^2}{\hat{\sigma}_t^2}$$

Multiplying and dividing the right-hand side by $\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 y_{t-1}^2$ we obtain

$$\hat{\alpha}_0 \sum_{t=2}^{T}\frac{1}{\hat{\sigma}_t^4} + \hat{\alpha}_1 \sum_{t=2}^{T}\frac{y_{t-1}^2}{\hat{\sigma}_t^4} = \sum_{t=2}^{T}\frac{y_t^2}{\hat{\sigma}_t^4}$$

and

$$\hat{\alpha}_0 \sum_{t=2}^{T}\frac{y_{t-1}^2}{\hat{\sigma}_t^4} + \hat{\alpha}_1 \sum_{t=2}^{T}\frac{y_{t-1}^4}{\hat{\sigma}_t^4} = \sum_{t=2}^{T}\frac{y_{t-1}^2 y_t^2}{\hat{\sigma}_t^4}$$

These two equations represent the ML estimator as the result of solving a system of equations which is the same system solved by the GLS estimator considered in subsection 4.2. Therefore, as ML and GLS are asymptotically equivalent, the effects of outliers on both estimators should be similar for large samples. Figure 2 plots the kernel estimates of the densities of the ML estimators of α_0 and α_1 when the ARCH(1) series are contaminated by a single outlier of size ω. This figure illustrates that when $\omega = 0$ the GLS and ML estimators are asymptotically equivalent. However, the finite sample distribution of both estimators can be rather different in the presence of large outliers and moderate sample sizes. Note that, for sample sizes of $T = 500$ and 1000 and outliers of sizes 10 and 15 standard deviations, the kernel estimated density of both $\hat{\alpha}_0^{\mathrm{ML}}$ and $\hat{\alpha}_1^{\mathrm{ML}}$ are bimodal and non-symmetric. The bimodality of the ML estimator in the presence of outliers has also been pointed out by Doornik and Ooms (2003). Hence, tests based on normality will be inadequate. Looking for example at the seventh row of Figure 2, we can see that in the presence of an outlier of size 15 standard deviations in a sample of size $T = 500$, $\hat{\alpha}_1^{ML}$ could take any value between 0 and 1, although values close to zero seem to be more probable, like what we had for $\hat{\alpha}_1^{OLS}$ and $\hat{\alpha}_1^{GLS}$. Finally, if the sample size is 5000, the sample distributions of the GLS and ML are similar. Therefore, it is important to point out that the results in Figure 2 suggest that, in moderate samples, the GLS estimator has certain advantages over the ML estimator in the presence of large isolated outliers. In particular, both estimators have similar negative biases but the dispersion of the GLS is smaller.

Figure 3 (column 3) plots the logarithm of the ratio of the sample variance and the estimated asymptotic variance averaged over all Monte Carlo replicates for $\hat{\alpha}_0^{ML}$ and $\hat{\alpha}_1^{ML}$ respectively. We can see that this ratio is larger than the ratio of the GLS estimator. Therefore, the asymptotic variance of the ML estimator underestimates, in the presence of large isolated outliers, the sample variance more than the GLS estimator.

The Monte Carlo densities when the series are contaminated by two consecutive outliers appear in Figure 4 (rows 3 and 7). As we can see in the plots, the effects caused by two consecutive outliers on the ML estimators are very similar to the effects caused by a single outlier. Finally, the effects of consecutive outliers on the estimated variances of the ML estimator are weaker than for the GLS estimator (see Figure 5).

4.4. *Quasi-Maximum Likelihood estimator*

As mentioned in the Introduction, the presence of outliers could be the reason of the excess kurtosis found in the standardized observations after an ARCH-type model has been fitted to explain the dynamic evolution of second order moments. However, many authors have claimed that this excess kurtosis can be explained by a heavy-tailed conditional distribution; see, for example, Bollerslev (1987), Baillie and Bollerslev (1989), Hsieh (1989), Nelson (1991) and Fiorentini *et al.* (2003) among many others. Furthermore, Sakata and White (1998) show that QML estimators based on heavy-tailed distributions are robust in the presence of outliers. Consequently, in this subsection, we study the finite sample behavior of the QML estimator based on maximizing the Student-likelihood when the data is generated by conditionally Gaussian ARCH(1) models contaminated by isolated and consecutive outliers. If no outliers are present, Newey and Steigerwald (1997) show that when the assumed and true densities are symmetric, the QML estimator is consistent and efficient.[2] The log-likelihood function is given by

$$L_S = \sum_{t=2}^{T}\left[\log\left(\Gamma\left(\frac{\eta+1}{2\eta}\right)\right) - \log\left(\Gamma\left(\frac{1}{2\eta}\right)\right) - \frac{1}{2}\left[\log\left(\frac{1-2\eta}{\eta}\right) + \log\pi \right.\right.$$
$$\left.\left. + \log(\alpha_0 + \alpha_1 y_{t-1}^2) + \frac{\eta+1}{\eta}\log\left(1 + \frac{\eta}{1-2\eta}\frac{{y_t}^2}{\alpha_o + \alpha_1 y_{t-1}^2}\right)\right]\right]$$

where $\Gamma(\cdot)$ is the gamma function and $\eta = 1/v$, where v are the degrees of freedom of the Student t distribution. The parameter η can be considered as a measure of tail thickness which always remains in the finite range $0 \leq \eta < 0.5$ if the conditional distribution is restricted to have finite variance, i.e. $v > 2$ (see Fiorentini *et al.*, 2003).

Rows 4 and 8 of Figure 2 plot the Monte Carlo densities of the QML estimator of α_0 and α_1 respectively, when the series are generated by the same ARCH(1) model considered before and contaminated by isolated outliers of size $\omega = 0$, 5, 10 and 15 marginal standard deviations. This figure shows that when $\omega = 0$, the behaviour of the QML estimator is very similar to ML. However, as postulated by Sakata and White (1998), maximizing the Student t likelihood protects against outliers even when they are rather large. It can be observed that the QML estimator is almost unaffected by the presence of outliers. Notice that even in small sample sizes the biases of $\hat{\alpha}_0^{\mathrm{QML}-t}$ and $\hat{\alpha}_1^{\mathrm{QML}-t}$ are almost negligible. The average over the Monte Carlo replicates of the degrees of freedom estimates is 83.64 when $\omega = 0$ and $T = 1000$ while it is 9.73 when $\omega = 15$. We obtain similar results for the other sample sizes. With respect to the estimated asymptotic standard deviations, the last column of Figure 3 shows that they are not affected by single outliers even if their sizes are large or the sample sizes are small. The same conclusions can be obtained from Figures 4 and 5 which plot the corresponding kernel densities of the QML estimators of α_0 and α_1 and the log-ratios of their variances obtained when the series are contaminated by two consecutive outliers.

5. EFFECTS OF OUTLIERS ON THE ESTIMATION OF GARCH MODELS

In this section we analyse the effects of outliers on three estimators of the parameters of GARCH(1,1) models. First, we analyse the finite sample properties of the ML and QML estimators described above for ARCH models. Second, we consider the closed-form estimator proposed by Kristensen and Linton (2006) which is based on the Yule–Walker equations corresponding to the ARMA representation of squared observations.

The robustness properties of the QML[3] estimator of the parameters of GARCH models have been analysed by Sakata and White (1998) and Mendes (2000). The former authors show that when the conditional mean is known, QML estimators based on thin-tailed distributions as the normal are not robust to outliers and have a breakdown point equal to zero. On the other hand, QML estimators obtained by maximizing a log-likelihood function with a fat-tailed distribution are resistant to outliers as long as there are no scale leverage points. Later, Mendes (2000) proves that the QML estimator obtained by maximizing the Gaussian log-likelihood has zero breakdown point and unbounded influence curves. In this article also, some Monte Carlo evidence is presented on the finite sample performance of the QML estimator in the presence of outliers concluding that the bias and standard errors of the QML estimates increase with the percentage of contamination and that for very large outliers, they ignore the dynamics in the second moments and act as an inflated scale estimate of independent observations. In this section, we generalize and extend this simulation evidence by carrying out detailed Monte Carlo experiments to analyse the biases caused by isolated and consecutive LO on the QML estimator of the parameters of GARCH(1,1) models.

Figure 6 contains the kernel estimates of the density of $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\hat{\beta}$ and $\hat{\alpha}_1 + \hat{\beta}$ for ML and QML estimates based on 1000 replicates, for a GARCH(1,1) model with parameters $\alpha_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta = 0.8$, contaminated with a single outlier of sizes $\omega = 0$, 5, 10 and 15 standard deviations. This figure shows that, unless the sample size is very large, like $T = 5000$, ML estimators as expected are not robust to the presence of outliers. The same conclusion is obtained by Mendes (2000) and Sakata and White (1998). The QML estimator has an interesting behaviour. It is robust for α_1, as in the case of ARCH models, but not for α_0, β and $\alpha_1 + \beta$. Note that the Student t-tails are robust for isolated outliers in ARCH models but for GARCH models one isolated outlier at time t affects the estimation of the conditional variance at time $t + 1$, and this variance will be used in the estimation of the conditional variance at time $t + 2$. Thus, an isolated outlier behaves as a patch of outliers for the estimation of the conditional variance. This explains the different behaviour of the QML estimator for ARCH and GARCH models.

In our Monte Carlo experiments we have also observed that although the generating model has finite fourth-order moment, the presence of outliers leads to a large proportion of ML estimates which do not satisfy the corresponding condition. For example, when $T = 500$ and $\omega = 15$, 36% of the ML estimates of α_1 and β do not satisfy it. In our simulations, we often observed replicates for which the estimated ML asymptotic covariance matrix is nearly singular. This could be due to the estimates $\hat{\alpha}_1^{\mathrm{ML}}$ and $\hat{\beta}^{\mathrm{ML}}$ taking values close to zero and one respectively, and the determinant of the information matrix being very close to zero. Sakata and White (1998) also observe in real data that, as a consequence of extreme outliers, the usual plug-in asymptotic covariance matrix could be nearly singular. Consequently, we do not report the results on the ratio of the sample variance and the estimated asymptotic variance although we have observed that, when the latter variances are defined, their behaviour is similar to the one observed for the ML estimator in Figures 3 and 5. Therefore, when using the estimated asymptotic ML covariance matrix we may have a false security on the inference on the GARCH parameters. The usual hypothesis testing methods will be highly unreliable.

Figure 7 plots kernel estimates of the density of the parameters for the same GARCH(1,1) model but now contaminated with two consecutive outliers. This figure shows that $\hat{\alpha}_0^{\mathrm{ML}}$ and $\hat{\alpha}_1^{\mathrm{ML}}$ overestimate the true parameters, and $\hat{\beta}^{\mathrm{ML}}$ is underestimating the true β. Note that if the outliers are large and the sample size is moderate, the sample densities of $\hat{\alpha}_1^{\mathrm{ML}}$ and $\hat{\beta}^{\mathrm{ML}}$ are such that standard inference is not reliable. Furthermore, this figure shows that large consecutive outliers can

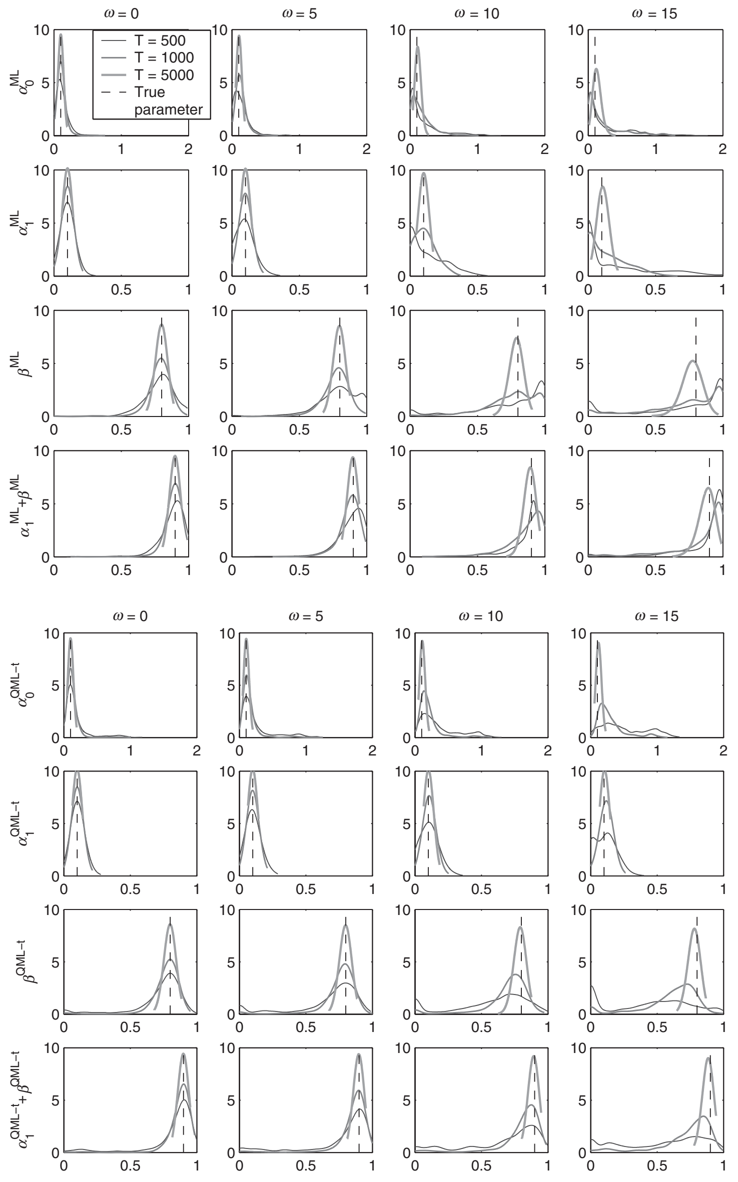

FIGURE 6. Kernel estimation of the density of ML and QML estimators of α_0 (rows 1 and 5), α_1 (rows 2 and 6), β (rows 3 and 7) and $\alpha_1 + \beta$ (rows 4 and 8) in a GARCH(1,1) model with a single outlier.

have dramatic effects on the estimated persistence. For example, when $\omega = 15$ and $T = 500$ or 1000, the estimated density of $\hat{\alpha}_1^{ML} + \hat{\beta}^{ML}$ has two modes, one around zero and the other close to one. The estimates of the persistence are only reliable for very large sample sizes. The biases caused by consecutive outliers on

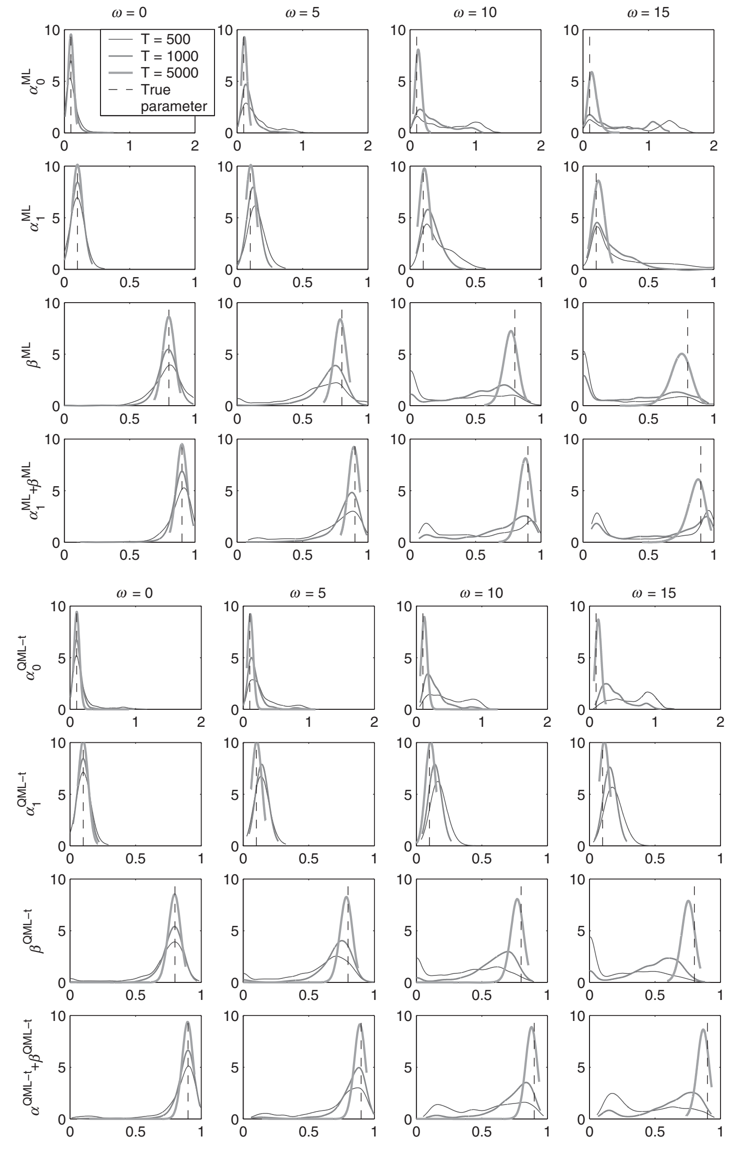

FIGURE 7. Kernel estimation of the density of ML and QML estimators of α_0 (rows 1 and 5), α_1 (rows 2 and 6), β (rows 3 and 7) and $\alpha_1 + \beta$ (rows 4 and 8) in a GARCH(1,1) model with two consecutive outliers.

the estimated asymptotic covariance matrix of the ML estimators are similar to those caused by a single outlier. Regarding the QML estimator, as before, it is robust for α_1 but not for α_0, β and $\alpha_1 + \beta$.

Finally, it is interesting to analyse the effects of outliers on the closed-form estimator for GARCH(1,1) models proposed by Kristensen and Linton (2006).

Assuming $y_t = \varepsilon_t \sigma_t$ where $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta \sigma_{t-1}^2$ we also have the ARMA(1,1) representation for the squared observations given by $y_t^2 = \alpha_0 + \phi y_{t-1}^2 + v_t - \beta v_{t-1}$, where $\phi = \alpha_1 + \beta$ and the noise, $v_t = \sigma_t^2(\varepsilon_t^2 - 1)$, is a zero-mean heteroscedastic uncorrelated sequence. Calling $r(h)$ as before the autocorrelation of the squared observations and using the relationship between these autocorrelations and the ARMA parameters, Kristensen and Linton (2006) propose the estimates

$$\hat{\phi} = \sum_{j=2}^{k} \frac{w_j r(j)}{r(j-1)}$$

where w_j is a weighting function and

$$\hat{\beta} = \frac{(b - \sqrt{b^2 - 4})}{2} \quad \text{with } b = \frac{(\hat{\phi}^2 + 1 - 2\,\hat{\phi} r(1))}{(\hat{\phi} - r(1))} \text{ and } \hat{\alpha}_1 = \hat{\phi} - \hat{\beta}.$$

In their simulation study, the weights are chosen as $w_j = 1/3$ for $j = 1,2,3$ and $w_j = 0$ for $j > 3$. Although, as shown by these authors, this estimate seems to work well without outliers, the effect of contamination in this estimate can be very large. We have seen that a single large outlier will make all the coefficients $r(j)$ small and thus the estimate $\hat{\phi}$, computed as a ratio of small numbers, will have a large variance and it may be unreliable. Note that the estimate $\hat{\phi}$ obtained does not need to be small. For patches of outliers, the coefficients $r(j)$ will be large and of similar size and the estimate $\hat{\phi}$ will be close to one and will not suffer for the large variability. However, the equation for b will not be reliable anymore, because of the large bias of the autocorrelation coefficients and b is often found to be smaller than 2. As this parameter is constrained to be $b > 2$ the censoring leads to $b = 2 + \varepsilon$ and then the estimate of $\hat{\beta}$ will be close to one and $\hat{\alpha}_1$ will be forced to be close to zero. This result has been checked by Monte Carlo.

6. CONCLUSIONS

Our results can be important in several directions. First of all, a lot of care should be taken when assuming conditional heteroscedasticity from a correlogram of squares with a large first-order autocorrelation followed by small coefficients for higher lags. We have seen that this pattern could be caused by two large consecutive outliers, regardless of whether the data-generating process is homoscedastic or heteroscedastic. Therefore, if the uncontaminated process is homoscedastic, after cleaning the outliers we will expect that the order-one autocorrelation will not be significant anymore. On the other hand, if the uncontaminated series is heteroscedastic, after correcting for outliers we expect that the first-order autocorrelation coefficient will decrease, while all the other coefficients will increase and become significant. Second, before rejecting the presence of conditional heteroscedasticity we should check that the series does not have large isolated outliers, as they bias all the autocorrelations towards zero.

Third, outliers may have a strong effect on the size and power of popular tests of conditional homoscedasticity based on autocorrelations of squares. On the one hand, a very large isolated outlier leads to tests which never reject the null hypothesis of homoscedasticity, regardless of whether the uncontaminated series is homoscedastic or heteroscedastic. On the other hand, two or more large consecutive outliers lead to tests which always reject the null hypothesis of homoscedasticity, even if the uncontaminated series is truly homoscedastic. Note that in the empirical analysis of financial returns it is more likely to find consecutive outliers because the series of returns is obtained as differences of the logarithmic prices. If the prices have a permanent shock, then the returns will show an isolated outlier but if the prices have a transitory movement during just one period of time, the returns will show two consecutive outliers. Therefore, care should be taken when the null of homoscedasticity is rejected especially if only the first-order autocorrelation is significantly different from zero.

Fourth, in moderately large samples, the QML estimator of ARCH models based on the Student likelihood, is more robust than the OLS, GLS and ML estimators in the presence of additive outliers. This QML estimator is robust against outliers without losing the good properties of ML for uncontaminated series. The OLS estimator has larger biases and the inference based on the estimated standard deviations is not reliable because they overestimate the true dispersion of the estimator. Furthermore, the negative estimates of the ARCH parameters that are sometimes obtained in empirical applications could be due to outliers. The sample distribution of the ML estimator of the ARCH parameters can be bimodal in the presence of outliers, implying important problems for inference.

Fifth, for GARCH (1,1) models the ML estimator of the parameters has very large dispersion even in moderately large samples as $T = 1000$, and the same happens for the GLS estimator. The QML estimator based on the Student likelihood is more robust than GLS and ML, but fails to be robust in general, especially for the estimation of the β parameter.

Thus, when fitting GARCH-type models to conditionally heteroscedastic series it is always advisable to check if the series is affected by outliers. There are several procedures proposed in the literature to test for outliers in the presence of GARCH effects and the comparison of these procedures will be the subject of further research.

ACKNOWLEDGEMENTS

Financial support from projects SEJ2004-03303, BEC2002.03720 and SEJ2005-02829/ECON by the Spanish Government and from Instituto Valenciano de

Investigaciones Económicas (IVIE) is gratefully acknowledged. We are also grateful to the comments of two anonymous referees that have helped us to improve this article. Any remaining errors are our own.

NOTES

1. Although the LM and $Q(m)$ tests are asymptotically equivalent, they test different null hypotheses. The null hypothesis of the LM test is that the parameters of an ARCH(p) model associated with lagged squared observations are equal to zero while the McLeod–Li test is model-free in the sense that the null hypothesis is that the first m autocorrelations are jointly equal to zero.
2. When this symmetry condition is not satisfied, Newey and Steigerwald (1997) show that, unless the conditional mean is identical to zero, the QML estimator is consistent if an additional location parameter is added to the model. Given that outliers can generate asymmetries, we also consider the introduction of this additional parameter. The modified ARCH(p) model is given by $y_t = (\delta + \varepsilon_t)\sigma_t$ where

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i}^2.$$

   The results are the same as those obtained when $\delta = 0$ and, consequently, they are not reported.
3. Note that in this article, the QML estimator refers to the estimator that maximizes the Student likelihood when the conditional distribution is not Student. However, the QML estimator is obtained when maximizing a likelihood that is not the same as the conditional distribution of the series. In fact, the most popular use of the QML estimator is when maximizing the Gaussian likelihood when the errors are not Gaussian.

Corresponding author; Esther Ruiz, Dpt. Estadística, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain. Tel.: 34 91 6249851; Fax: 34 91 6249849; E-mail: ortega@est-econ.uc3m.es

REFERENCES

Aggarwal, R., Inclan, C. and Leal, R. (1999) Volatility in emerging stock markets. *Journal of Financial and Quantitative Analysis* 34, 33–5.

Baillie, R. and Bollerslev, T. (1989) The message in daily exchange rates: a conditional-variance tale. *Journal of Business & Economic Statistics* 7, 297–305.

Balke, N. S. and Fomby, T. B. (1994) Large shocks, small shocks and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics* 9, 181–200.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–27.

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69, 542–7.

Bollerslev, T, Engle, R.F. and Nelson, D.B. (1994) ARCH models. In *Handbook of Econometrics, Vol. 4* (eds R. F. Engle and D. L. McFadden). New York: Elsevier Science, 2961–3033.

Bose, A. and Mukherjee, K. (2003) Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analysis* 24, 127–36.

Carnero, M. A (2003) Heterocedasticidad condicional, Atípicos y Cambios de nivel en series temporales financieras. Ph.D. Thesis, Universidad Carlos III de Madrid.

Doornik, J. A. and Ooms, M. (2003) Multimodality in the GARCH regression model. Manuscript, Nuffield College, University of Oxford.

Engle, R. F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Fiorentini, G., Sentana, E. and Calzolari, G. (2003) Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student *t* innovations. *Journal of Business & Economic Statistics* 21, 532–46.

Franses, P. H. and Ghijsels, E. (1999) Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting* 15, 1–9.

Franses, P. H., Van Dijk, D. and Lucas, A. (2004) Short patches of outliers, ARCH and volatility modelling. *Applied Financial Economics* 14, 221–31.

Hsieh, D. (1989) Modelling heteroscedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics* 7, 307–17.

Hotta, L. K. and Tsay, R. S. (1998) Outliers in GARCH processes. Manuscript. Graduate School of Business, University of Chicago.

Kristensen, D. and Linton, O. (2006) A closed-form estimator for the GARCH(1,1)-model. *Econometric Theory* 22, 323–37.

Lee, J. H. H. and King, M. L. (1993) A locally most mean powerful based score test for ARCH and GARCH regression disturbances. *Journal of Business and Economic Statistics* 11, 17–27.

Li, J. and Kao, C. (2002) A bounded influence estimation and outlier detection for ARCG/GARCH models with an application to foreign exchange rates. Manuscript. Finance and Insurance group, Northeastern University.

Ling, S. and McAleer, M. (2003) Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* 19, 280–310.

McLeod, A. J. and Li, W. K. (1983) Diagnostic checking ARMA time series models using squared-residual correlations. *Journal of Time Series Analysis* 4, 269–73.

Mendes, B. V. M. (2000) Assessing the bias of maximum likelihood estimates of contaminated GARCH models. *Journal of Statistical Computation and Simulation* 67, 359–76.

Muller, N. and Yohai, V. J. (2002) Robust estimates for ARCH processes. *Journal of Time Series Analysis* 23, 341–75.

Nelson, D. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.

Newey, W. K. and Steigerwald, D. G. (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–99.

Peña, D. (2001) Outliers, influential observations and missing data. In *A Course in Time Series* (eds D. Peña, G. C. Tiao and R. S. Tsay). New York: Wiley, 136–70.

Peña, D. and Rodriguez, J. (2002) A powerful portmanteau test of lack of fit in time series. *Journal of the American Statistical Association* 97, 601–10.

Rodriguez, J. and Ruiz, E. (2005) A powerful test for conditional heteroscedasticity for financial time series with highly persistent volatilites. *Statistica Sinica* 15, 505–25.

Sakata, S. and White, H. (1998) High breakdown point conditional dispersion estimation with application to S&P500 daily returns volatility. *Econometrica* 66, 529–67.

Van Dijk, D., Franses, P. H. and Lucas, A. (1999) Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* 14, 539–62.

Weiss, A. (1986) Asymptotic theory for ARCH models. *Econometric Theory* 2, 107–31.

Wooldridge, J. M. (1990) A unified approach to robust regression based specification tests. *Econometric Theory* 6, 17–43.