Outliers and misleading leverage effect in asymmetric GARCH-type models disposition effect

M. Angeles Carnero and Ana Pérez

Ivie

Outliers and misleading leverage effect in asymmetric GARCH-type models[*]

M. Angeles Carnero and Ana Pérez[**]

Abstract

This paper illustrates how outliers can affect both the estimation and testing of leverage effect by focusing on the TGARCH model. Three estimation methods are compared through Monte Carlo experiments: Gaussian Quasi-Maximum Likelihood, Quasi-Maximum Likelihood based on the t-Student likelihood and Least Absolute Deviation method. The empirical behavior of the t-ratio and the Likelihood Ratio tests for the significance of the leverage parameter is also analyzed. Our results put forward the unreliability of Gaussian Quasi-Maximum Likelihood methods in the presence of outliers. In particular, we show that one isolated outlier could hide true leverage effect whereas two consecutive outliers bias the estimated leverage coefficient in a direction that crucially depends on the sign of the first outlier and could lead to wrongly reject the null of no leverage effect or to estimate asymmetries of the wrong sign. By contrast, we highlight the good performance of the robust estimators in the presence of an isolated outlier. However, when there are patches of outliers, our findings suggest that the sizes and powers of the tests as well as the estimated parameters based on robust methods may still be distorted in some cases. We illustrate these results with two series of daily returns, namely the Spain IGBM Consumer Goods index and the futures contracts of the Natural gas.

Keywords: Conditional heteroscedasticity, QMLE, Robust estimators, TGARCH, AVGARCH.

JEL classification numbers: C22, G10, Q40.

[**] M.A. Carnero (corresponding autor): Dpto. Fundamentos del Análisis Económico, Universidad de Alicante. Campus de San Vicente del Raspeig, Alicante. Spain. Tel: +34 965 903400. E-mail: acarnero@ua.es. A. Pérez: Dpto. Economía Aplicada and IMUVA, Universidad de Valladolid. Spain. E-mail: perezesp@eaee.uva.es.

1 Introduction

One of the empirical stylized facts about the series of financial returns is the leverage effect, which refers to the asymmetric response of volatility to positive and negative past returns. In particular, the volatility tends to be higher following negative return shocks ('bad' news) than following positive shocks ('good' news) of the same magnitude. This feature conveys a generally negative cross-correlation between lagged asset returns and volatility; see Black (1976) who originally put it forward using the debt-to-equity ratio and Engle (2011) who provides the economic underpinning of volatility asymmetries following simple asset pricing theory. See also Hibbert *et al.* (2008) for a behavioral explanation of the negative asymmetric return–volatility relationship.

In the econometric literature there have been several methods proposed to represent and estimate conditional heteroscedasticity and leverage effect, like the asymmetric GARCH-type models (see the review in Rodríguez and Ruiz (2012)), the non-parametric high-frequency methods in Andersen *et al.* (2001) and the observation-driven models proposed by Creal *et al.* (2013) and Harvey (2013), to name but a few. In practice, when these methods are applied to estimate and test for the leverage effect, mixed results come up. For example, Andersen *et al.* (2001) found statistically significant leverage effect for most of the stocks returns of the DJIA stock market index by using realized volatilities, although they point out that this effect has a marginal economic importance. In turn, Zivot (2009) and Rodríguez and Ruiz (2012) also report significant leverage effect by applying GARCH-type methods to different series of returns, including a particular asset stock, a stock market index and exchange rates. However, Ait-Sahalia *et al.* (2013) discuss what they call *the leverage effect puzzle*, which relies on the fact that the empirical correlation between returns and changes in volatility estimated from high frequency data becomes nearly zero for most assets tested, despite the many economic reasons for expecting such correlation to be negative. In Energy markets, it is also common to find different results concerning the leverage effect. For example, Kristoufek (2014) finds the inverse leverage effect (positive correlation between returns and volatility) in future prices of the natural gas by using correlation-based methods for non-stationary series, whereas Chkili *et al.* (2014) find the standard leverage effect when asymmetric and long-memory models are estimated to spot and future returns of the same commodity.

4

The mixed results on the sign and the significance of the leverage effect could be partly explained, among other reasons, by the harmful effect of the extreme observations usually encountered in the returns. For instance, it is already well known that the presence of outliers can have misleading effects on the identification of both conditional heteroscedasticity and leverage effect; see Carnero *et al.* (2007) and Carnero *et al.* (2016), respectively. Moreover, it is also proved that the presence of outliers renders the Gaussian Quasi Maximum Likelihood (QML) estimators unreliable in symmetric GARCH models; see Sakata and White (1998), Mendes (2000), Carnero *et al.* (2007) and Muler and Yohai (2008), among others. In this context, robust estimators based on maximizing non-Gaussian heavy-tailed likelihoods have been proposed and their main properties have been established; see, for instance, Newey and Steigerwald (1997), Sakata and White (1998), Berkes and Horvath (2004) and Fan *et al* (2014). Another robust alternative for GARCH-type models is the log-transform-based Least Absolute Deviations (LAD) estimator proposed by Peng and Yao (2003) and further discussed in Huang *et al.* (2008). Related works also include Pan *et al.* (2008), Francq and Zakoian (2013) and Hill (2015), among others. Alternatively, some authors deal with this problem by applying methodologies based on detecting and correcting outliers; see, for example, Laurent *et al.* (2016) and the references therein.

In this paper, we face the problem of how outliers can affect the inference on the leverage parameter in asymmetric GARCH-type models: Does the sign of the outliers matter? How much the size of the outliers worsen the results? Is the effect of one isolated outlier comparable to that of consecutive outliers? Which estimation method is more resistant to outliers? To answer these questions, we conduct an extensive Monte Carlo study that includes models with high, low and none leverage and different types of outliers (isolated and in patches, positive and negative, big and small). For each setting, we compare the performance of three estimation methods: QML, Quasi-Maximum Likelihood estimator based on maximizing the Student-t likelihood (QML-t) and LAD. We also analyze the size and power of the t-ratio and the Likelihood Ratio tests for the significance of the leverage parameter. Among the several asymmetric GARCH-type models proposed in the literature, we focus on the Threshold GARCH (TGARCH) model of Zakoian (1994) since, according to Rodríguez and Ruiz (2012), this is more flexible than its competitors to properly represent the dynamics of financial returns.

Previous works related to our paper are Pan *et al.* (2008), who establish the asymptotic properties of both QML and LAD estimators for a general model that nests the TGARCH and compare, for a particular parametrization, the finite sample performance of both estimators, and Francq and Zakoian (2013), who compare the asymptotic relative efficiencies of QML and LAD estimators for a model that also nests the TGARCH. Our paper provides new insights on the topic in two ways. First, we also evaluate the QML-t estimator, that turns out to be more robust than LAD in most cases, and second, our Monte Carlo study is exhaustive enough to cover the multiple problems often found in real data where different type of outliers may come up.

Our results show that QML-based methods become unreliable in the presence of outliers and could lead to either hide true leverage effects (when there is one isolated outlier) or detect spurious leverage or leverage of the wrong sign (in the presence of two consecutive outliers). As expected, the robust estimators considered (QML-t and LAD) always outperform QML although they are still slightly biased in the presence of big consecutive outliers. In this case, the bias direction crucially depends on the sign of the first outlier and could lead to wrongly reject the null of no leverage effect. These results are further enhanced in our empirical application, where two particular series of financial returns are analyzed, namely a daily series of the Spain IGBM Consumer Goods index including one isolated negative outlier and a daily series of futures contracts of Natural gas including two consecutive outliers of opposite sign being the first one positive.

The rest of the paper is organized as follows. Section 2 reviews the TGARCH model and describes the three estimation methods to be analyzed (QML, QML-t and LAD) and the two significance tests considered (t-ratio and Likelihood Ratio tests). Section 3 is devoted to the finite sample performance of these estimators and tests in the presence of outliers for different parameter sets and different types of outliers. Section 4 illustrates our main results with an empirical application based on the two series of daily returns mentioned above. Finally, Section 5 concludes the paper with a summary of the main conclusions.

2 Statistical inference in the TGARCH model

2.1 TGARCH model: definition and main properties

As Rodríguez and Ruiz (2012) point out, the TGARCH model is an appropriate and flexible GARCH-type model to represent the features and dynamic properties of financial returns, namely excess kurtosis, conditional heteroscedasticity and leverage effect. In this model, the series of demeaned returns, y_t, is specified by the following equation:

$$y_t = \sigma_t \, \varepsilon_t, \tag{1}$$

where σ_t is the volatility and ε_t is a sequence of independent and identically distributed random variables with zero mean and unit variance. To accommodate the asymmetric relationship between past returns and volatility, σ_t is parametrized as a function of both the magnitude and the sign of past returns. In particular, the equation for the volatility in the TGARCH(1,1) model is[1]

$$\sigma_t = \alpha_0 + \alpha \, |y_{t-1}| \, + \beta \sigma_{t-1} + \delta y_{t-1}. \tag{2}$$

When y_{t-1} is positive, the volatility response is linear in y_{t-1} with slope $(\delta + \alpha)$ but if y_{t-1} is negative, the slope of the response is $(\delta - \alpha)$. Thus, the volatility can respond asymmetrically to rises and falls in stock prices and the value of δ is expected to be negative. Under the constraints $\alpha_0 > 0$, $\beta \geq 0$ and $\alpha \geq |\delta|$, σ_t is always positive and represents the conditional standard deviation of y_t. Moreover, the model is covariance stationary if $\delta^2 < 1 - \alpha^2 - \beta^2 - 2\alpha\beta\nu_1$, where $\nu_1 = E|\varepsilon_t|$. When this condition is satisfied, the marginal variance of returns is given by

$$\sigma_y^2 = \alpha_0^2 \frac{1 + \alpha\nu_1 + \beta}{(1 - \alpha\nu_1 - \beta)(1 - \delta^2 - \alpha^2 - \beta^2 - 2\alpha\beta\nu_1)}.$$

An advantage of the TGARCH model is that it parametrizes the conditional standard deviation rather than the conditional variance. This makes it easier to work out analytical expressions for its unconditional higher-order moments and cross-moments. He and Teräsvirta (1999) and He *et al.* (2008) provide the moment structure as well

---

[1]It is important to be aware that other parametrizations are possible (see Appendix A) and that the GJR model, proposed by Glosten et al. (1993), is sometimes erroneously referred to as TGARCH.

as the autocorrelation of the squared and absolute returns and the cross-correlation between squared and lagged returns; see also the results in Hwang and Basawa (2004). Moreover, as the TGARCH model involves absolute values rather than squares, it is expected to be less sensitive to extreme observations than similar models involving conditional variances and squared returns, like the GJR and the QARCH models in Glosten *et al.* (1993) and Sentana (1995), respectively.

Another advantage of the TGARCH model is that it contains, as a special case, a model without leverage whose properties are very similar to those of the standard GARCH. Such a model is the absolute-value GARCH (AVGARCH) model of Taylor (1986) and Schwert (1989), that comes up by taking $\delta = 0$ in (2). Thus, model selection can be easily performed by testing the significance of the leverage parameter δ with the usual t-ratio test and/or the Likelihood Ratio test. On the other hand, the TGARCH model is a particular case of the Asymmetric Power ARCH (A-PARCH) model proposed by Ding *et al* (1993) with power parameter equals to one. Hence, the asymptotic theory in Pan *et al.* (2008) and Francq and Zakoian (2013) regarding estimation and hypothesis testing in that general model can be applied. By contrast, the asymptotic theory for other popular asymmetric GARCH-type models, like the EGARCH model proposed by Nelson (1991), is scarce. Only some results on the consistency of the Gaussian QML estimator in that model can be found in Straumann and Mikosch (2006) and Wintenberger (2013).

2.2 Gaussian QML estimation

The TGARCH(1,1) model defined in equations (1)-(2) can be estimated by maximizing the conditional log-likelihood function, which, given initial values y_0 and σ_0, is as follows

$$L(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} \left\{ -\frac{1}{2} \log \sigma_t^2 + \log f\left(\frac{y_t}{\sigma_t}\right) \right\}, \tag{3}$$

where $\theta = (\alpha_0, \alpha, \beta, \delta)'$ denotes the parameter vector to be estimated and $f(\cdot)$ is the probability density of ε_t. In particular, if the log-likelihood function in (3) is computed as if ε_t were assumed to be $N(0,1)$, the corresponding Gaussian log-likelihood function,

denoted as $L_G(\theta)$, comes up, namely

$$L_G(\theta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\left(\log\sigma_t^2 + \frac{y_t^2}{\sigma_t^2}\right). \tag{4}$$

The resultant estimator obtained from maximizing $L_G(\theta)$ is the well-known QML estimator. This estimator, that will be denoted as $\widehat{\theta}_{QML}$, is the most commonly used one for GARCH-type models and, in particular, for the TGARCH model introduced above; see, for instance, Zivot (2009) and Francq and Zakoian (2010, chapter 10). Moreover, this estimator is provided in most software packages, such as E-VIEWS, G@RCH4.0, MFE MATLAB Toolbox, SAS, Stata, Splus and R. Obviously, if the true distribution of ε_t is $N(0,1)$, the resultant estimator will be the Maximum Likelihood estimator.

Pan *et al.* (2008) show that QML is consistent and asymptotically normal for an asymmetric power-transformed GARCH model that includes as a particular case the TGARCH model, provided that ε_t is symmetrically distributed with $E\varepsilon_t^2 = 1$ and $E\varepsilon_t^4 < \infty$ and some regularity assumptions hold. In such a framework, they show that

$$\sqrt{T}(\widehat{\theta}_{QML} - \theta) \xrightarrow{\mathcal{L}} N(0, \Lambda^{-1}\Omega\Lambda^{-1}), \tag{5}$$

where

$$\Lambda(\theta) = E\left(\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right) \text{ and } \Omega(\theta) = E\left(\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}\right).$$

Hence, we can approximate the asymptotic variance of $\widehat{\theta}_{QML}$ by the so-called "sandwich" estimator

$$Var(\widehat{\theta}_{QML}) \approx H(\widehat{\theta}_{QML})^{-1}B(\widehat{\theta}_{QML})H(\widehat{\theta}_{QML})^{-1}, \tag{6}$$

where $H(\theta)$ denotes the Hessian matrix of the log-likelihood and $B(\theta)$ is the inner product of the gradient (or score) of the log-likelihood, namely

$$H(\theta) = \sum_{t=1}^{T}\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}, \quad B(\theta) = \sum_{t=1}^{T}\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}.$$

On the other hand, Muler and Yohai (2008) point out that maximizing (4) is equivalent to minimizing

$$M_{0,T}(\theta) = \frac{1}{T-2}\sum_{t=2}^{T}\rho_0(x_t - \log\sigma_t^2(\theta)),$$

where $x_t = \log(y_t^2)$ and $\rho_0 = -\log(g_0)$, where g_0 is the density of $\log(\varepsilon_t^2)$ when the distribution of ε_t is assumed to be N(0,1), that is

$$g_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(e^x - x)},$$

so that

$$\rho_0(x) = \log(\sqrt{2\pi}) + \frac{1}{2}(e^x - x). \tag{7}$$

Hence, the QML estimator, that could be alternatively defined as

$$\widehat{\theta}_{QML} = \arg\min_{\theta} M_{0,T}(\theta),$$

corresponds to an M-estimator with auxiliary function ρ_0 in (7). This function as well as its 1st derivative, $\rho_0' = \frac{1}{2}(e^x - 1)$, are both unbounded, rendering QML not robust, i.e., a few outliers can have a large influence on this estimator.

The lack of robustness of the QML estimator in symmetric GARCH models is already well documented; see Sakata and White (1998), Mendes (2000), Carnero *et al.* (2007) and Muler and Yohai (2008), among others. In Section 3 we analyze the effect of the outliers on QML in TGARCH models through Monte Carlo experiments. In particular, we investigate if the sign of the outliers, that is irrelevant in symmetric models, makes any difference when estimating the parameters of asymmetric models.

2.3   QML-t estimation

In order to gain resistance against outliers, some authors propose to use estimators based on maximizing non-Gaussian heavy-tailed log-likelihoods; see, for instance, Sakata and White (1998). Actually, in the seminal paper of Nelson (1991), the EGARCH model is estimated by maximizing the loglikelihood function in (3) assuming that ε_t follows a GED distribution normalized to have zero mean and unit variance. Another common practise is to maximize the conditional log-likelihood in (3) computed as if ε_t followed a Student-t distribution with υ degrees of freedom normalized to have zero mean and unit variance. In such a case, the log-likelihood function, denoted as L_{Stud}, becomes:

$$L_{Stud}(\theta) = T \log\left(\frac{\Gamma((\upsilon+1)/2)}{\sqrt{\pi(\upsilon-2)}\Gamma(\upsilon/2)}\right) - \frac{1}{2}\sum_{t=1}^{T}\left[\log\sigma_t^2 + (\upsilon+1)\log\left(1 + \frac{1}{\upsilon-2}\frac{y_t^2}{\sigma_t^2}\right)\right], \tag{8}$$

where the vector parameter is $\theta = (\alpha_0, \alpha, \beta, \delta, \upsilon)'$. The resultant estimator obtained from maximizing $L_{Stud}(\theta)$ is the so-called QML-t estimator denoted as $\widehat{\theta}_{QML-t}$.

As far as we know, no asymptotic theory exists for QML-t estimation in the context of asymmetric GARCH models. Hence, we will assume the usual practice of researchers using GARCH-type models that the asymptotic distribution of QML-t is that in (5) and we will approximate its asymptotic variance by the "sandwich" estimator in (6) replacing $\widehat{\theta}_{QML}$ by $\widehat{\theta}_{QML-t}$. Empirical applications performing QML-t estimation of the TGARCH model can be found in Zivot (2009), Francq and Zakoian (2010, chapter 10) and Rodríguez and Ruiz (2012).

Muler and Yohai (2008) point out that maximizing (8) is equivalent to minimizing

$$M_{1,T}(\theta) = \frac{1}{T-2}\sum_{t=2}^{T} \rho_{1,\upsilon}(x_t - \log \sigma_t^2(\theta)),$$

where $x_t = \log(y_t^2)$ and $\rho_{1,\upsilon} = -\log(g_1)$, where g_1 is the density of $\log(\varepsilon_t^2)$ when the distribution of ε_t is assumed to be a standardized Student, i.e.,

$$g_1(x) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\pi(\upsilon-2)}\Gamma\left(\frac{\upsilon}{2}\right)}e^{\frac{x}{2}}\left(1 + \frac{e^x}{\upsilon-2}\right)^{-\frac{\upsilon+1}{2}},$$

so that

$$\rho_{1,\upsilon}(x) = -\frac{x}{2} + \frac{\upsilon+1}{2}\log\left(1 + \frac{e^x}{\upsilon-2}\right) - \log\Gamma\left(\frac{\upsilon+1}{2}\right) + \log\left(\sqrt{\pi(\upsilon-2)}\Gamma(\frac{\upsilon}{2})\right). \quad (9)$$

Hence, the QML-t estimator could be alternatively defined as

$$\widehat{\theta}_{QML-t} = \arg\min_{\theta} M_{1,T}(\theta),$$

and so it also corresponds to an M-estimator with the auxiliary function $\rho_{1,\upsilon}$ in (9). This function is unbounded but its 1st derivative, $\rho'_{1,\upsilon}(x) = \frac{\upsilon e^x - \upsilon + 2}{2(\upsilon - 2 + e^x)}$, is bounded; thus, QML-t is expected to be more robust than QML, although it can still be affected by some type of outliers.

Carnero *et al.* (2007) analyze the robustness of QML-t for symmetric ARCH and GARCH models and conclude that it is resistant against outliers without loosing efficiency when estimating the parameters of ARCH models. However, it fails to be robust when estimating the parameter β of the GARCH(1,1) model. In Section 3 we analyze the finite sample behavior of the QML-t estimator, as compared to QML, in TGARCH models in the presence of outliers.

2.4 LAD estimation

As an alternative to QML, Peng and Yao (2003) propose the LAD estimator and they show that, in the context of GARCH models, this estimator is robust to heavy tails of the innovation distribution and it is asymptotically normal and unbiased under mild conditions for the error distribution. Huang *et al.* (2008) perform a comparison between both QML and LAD, where the latter is viewed as a quasi-maximum likelihood estimator based on the hypothesis that the log-squared innovations follow a Laplace distribution. Pan *et al.* (2008) establish the asymptotic properties of both QML and LAD estimators for an asymmetric power-transformed GARCH model that nests the TGARCH. They also compare, for a particular parametrization, the finite sample performance of both estimators and show that the LAD is more accurate than QML for heavy-tailed errors.

For the LAD estimator to be applied, the model should be reparametrized in such a way that the median (instead of the mean) of the squared innovations is equal to 1, while the mean of the innovations remains unchanged and equal to 0. In particular, for the TGARCH model we are interested in, the reparametrization is as follows. Let $M = median(\varepsilon_t^2) > 0$ and let $\sigma_t^* = M^{1/2}\sigma_t$ and $\varepsilon_t^* = M^{-1/2}\varepsilon_t$, so that $median(\varepsilon_t^{*2}) = 1$. Then, the TGARCH model defined in equations (1)-(2) may now be expressed as

$$
\begin{aligned}
y_t &= \sigma_t^* \, \varepsilon_t^*, \\
\sigma_t^* &= \alpha_0^* + \alpha^*|y_{t-1}| + \beta\sigma_{t-1}^* + \delta^* y_{t-1},
\end{aligned}
$$

where $\alpha_0^* = M^{1/2}\alpha_0$, $\alpha^* = M^{1/2}\alpha$, $\delta^* = M^{1/2}\delta$ and the parameter vector to be estimated is $\theta^* = (\alpha_0^*, \alpha^*, \beta, \delta^*)'$. Notice that, under this new parametrization, the parameters α_0^*, α^* and δ^* differ from those in the old setting by a common positive constant factor while the parameter β remains unchanged. The LAD estimator is based on the regression equation for the log-squared transformation, namely

$$
\log y_t^2 = \log \sigma_t^{*2} + \log \varepsilon_t^{*2},
$$

where the error terms $\log \varepsilon_t^{*2}$ are independent and identically distributed with median equals to 0. Now, the LAD estimator of θ^* is defined as:

$$
\widehat{\theta}^*{}_{LAD} = \arg\min_{\theta} \sum_{t=2}^{T} \left| \log y_t^2 - \log \sigma_t^{*2}(\theta) \right| .
$$

Notice that, since some of the parameters estimated by LAD differ from the original parameters by a common positive factor, to compare the LAD estimates to those obtained by other methods, the former should be corrected by the corresponding scale factor $M^{-1/2}$. Moreover, for the standard errors of the LAD estimators to be computed, it is required to estimate the density of the log-squared innovations by the kernel method. Hence, the additional problem of choosing the bandwidth and the kernel function to be used should be addressed; see Pan *et al.* (2008).

Interestingly, Muler and Yohai (2008) show that the LAD estimator can also be regarded as an M-estimator defined as

$$\widehat{\theta}^*_{LAD} = \arg\min_{\theta} M_{2,T}(\theta),$$

with

$$M_{2,T}(\theta) = \frac{1}{T-2}\sum_{t=2}^{T}\rho_2(x_t - \log\sigma_t^2(\theta)),$$

where $x_t = \log(y_t^2)$ and ρ_2 is the following auxiliary function

$$\rho_2(x) = |x - u_0|, \tag{10}$$

where $u_0 = \log M$. The function ρ_2 in (10) is unbounded but $\rho_2'(x) = sign(x - u_0)$ is bounded. Thus, as QML-t, this estimator is more robust than QML, although large outliers can still have a strong effect on it, as we will see in Section 3.

To summarize, Figure 1 plots the functions ρ_0, $\rho_{1,v}$ and ρ_2 in (7), (9) and (10), respectively, as well as their 1st derivatives. For the QML-t, we consider $v = 3$ and for the LAD we take $u_0 = 0$. This figure clearly shows up the differences and similarities between the three functions: all of them are unbounded but, whereas ρ_0' is still unbounded, $\rho_{1,v}'$ and ρ_2' are both bounded, making the corresponding estimators, QML-t and LAD, more robust to outliers than QML.

2.5   Testing the significance of the leverage coefficient

The goal of estimating asymmetric GARCH models, like the TGARCH, is capturing the leverage effect through the leverage parameter δ. Hence, once the model has been

Figure 1: Function ρ and its first derivative in M-estimators
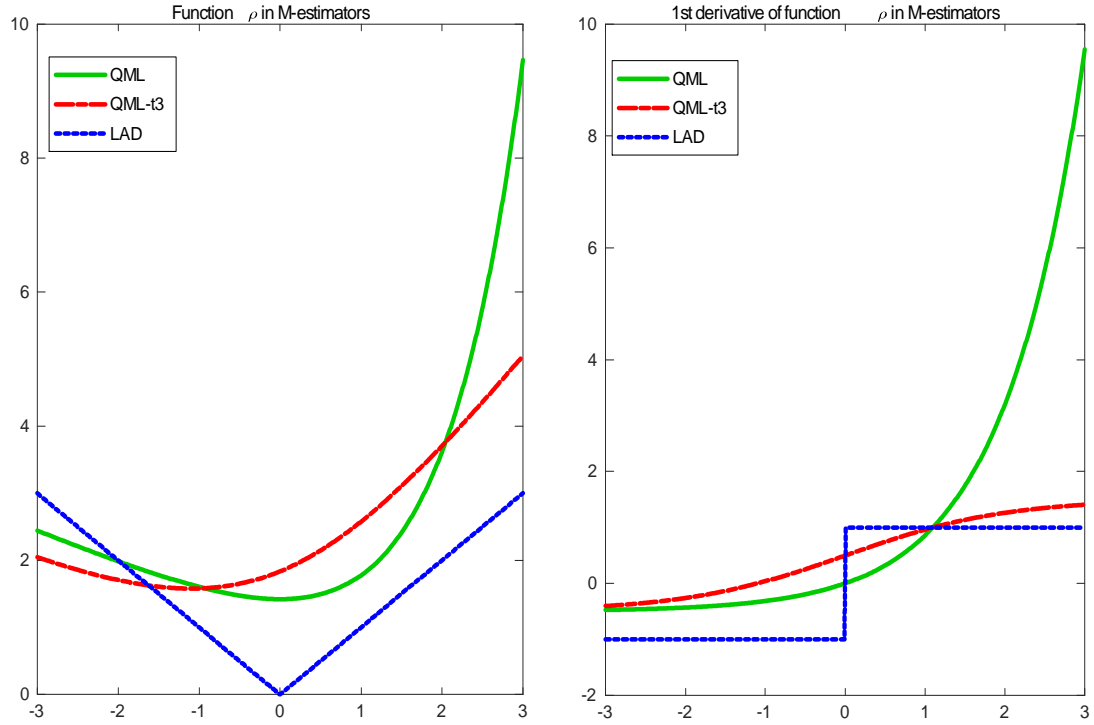
estimated, the following significance testing problem is usually considered

$$H_0 : \delta = 0 \ \ \text{against} \ \ H_1 : \delta \neq 0.$$

The most widely used test for this problem is the t test based on the QML estimation of the model. This test employs the statistic $\mathbf{t} = \widehat{\delta}/(s.e.(\widehat{\delta}))$, where the standard error is computed as the square root of the corresponding diagonal element of the approximate asymptotic variance (6); see, for instance, Zivot (2009). At the asymptotic significance level α, the standard rejection region is thus

$$\{|\mathbf{t}| > \Phi^{-1}(1 - \alpha/2)\}. \tag{11}$$

Alternatively, the Likelihood Ratio (LR) test can be performed. This test employs the statistic $LR = -2(L_c - L_u)$, where L_c is the constrained log-likelihood (estimating the model under the null) and L_u is the unconstrained log-likelihood. At the asymptotic significance level α, the standard rejection region of the LR test is

$$\{LR > \chi_1^2(1 - \alpha)\}, \tag{12}$$

where $\chi_1^2(1 - \alpha)$ is the $(1 - \alpha)-$quantile of the χ_1^2 distribution. Other approaches are possible; see, for instance, the Wald test based on LAD estimation in Pan *et al.* (2008).

Despite the popularity of the LR and t tests, Francq and Zakoian (2010) warn about using their standard forms in (12) and (11), respectively, to test for the significance of GARCH coefficients. In particular, they point out that the standard version of the t-test in (11) is not of asymptotic level α but only $\alpha/2$. In turn, they propose modified versions of both tests which are appropriate for statistical inference in symmetric GARCH models. Whether those results apply to asymmetric GARCH models, like the TGARCH, is still an open question, but we conjecture that this would be the case. Furthermore, the problem will be enhanced in the presence of outliers, as we will see in the next section, where we analyze the empirical behavior of the standard versions of the t and LR tests in (11) and (12), respectively, in the context of TGARCH models contaminated with outliers.

3   Monte Carlo simulation

In this section we report the results on several Monte Carlo experiments comparing the performance of the three estimators described in Section 2, for two TGARCH models and a symmetric AVGARCH model contaminated with one isolated outlier and with two consecutive outliers of different sizes and sign. As far as we know, this is the first time that these three estimators are compared in such a framework. Previous work on standard GARCH models include Peng and Yao (2003), who compare QML and LAD in ARCH(2) and GARCH(1,1) models, and Muler and Yohai (2008), who compare, in the GARCH(1,1) model, the behavior of several estimators, including QML, QML-t (with fixed degrees of freedom $v = 3$) and LAD. Also, Pan *et al.* (2008) compare numerically QML and LAD for a general PTTGARCH(1,1) model that allows for leverage effect. However, their comparison is based on the average absolute error of all the model parameters, a criteria that misses the sign of the leverage parameter and does not allow to disentangle the effect of outliers on each parameter.

3.1   Data generation and estimators

In our Monte Carlo experiments, the data are generated by the following scheme:

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

with $\varepsilon_t \sim N(0,1)$, so that $E\varepsilon_t^2 = 1$ and $M = median(\varepsilon_t^2) = 0.454936$. The volatility process σ_t is generated by equation (2) with three true parameter sets:

Parameter set 1	$\alpha_0 = 0.0475,\ \alpha = 0.15,\ \beta = 0.83,\ \delta = -0.05$
Parameter set 2	$\alpha_0 = 0.0746,\ \alpha = 0.12,\ \beta = 0.825,\ \delta = -0.11$
Parameter set 3	$\alpha_0 = 0.0746,\ \alpha = 0.12,\ \beta = 0.825,\ \delta = 0$

The first two parameter sets define TGARCH models capturing leverage effects of different magnitude, while the latter defines a symmetric AVGARCH (without leverage) that is nested in the TGARCH model with parameter set 2. This allows us to evaluate the properties of some statistical tests for model selection, such as the Likelihood Ratio

test. Moreover, the parameters values of the two TGARCH models have been chosen to resemble the values usually encountered in real empirical applications and to make them somehow comparable, since their marginal variance and kurtosis are very similar. In particular, with these parameter sets, the marginal variance, the kurtosis and the 1st-order cross-correlation between squared and past returns are the following:

	Marginal variance	Kurtosis	1st-order cross-correlation
Parameter set 1	1	5.389	-0.0647
Parameter set 2	1	5.727	-0.135
Parameter set 3	0.9175	3.492	0

Since we are interested in the effect of both the size and the sign of the outliers, we consider positive and negative outliers and we take the case of no outliers as a benchmark. In particular, the sizes of the outliers considered are the following: $\omega = \{0, \pm 10, \pm 20, \pm 30, \pm 40, \pm 50\}$. For each model, we generate 1000 independent samples of size $T = 1000$.[2] Then we contaminate each sample, first, with $k = 1$ isolated outlier of size ω at time $t = T/2$ and second, with $k = 2$ consecutive outliers of the same size but opposite signs placed at $t = \{T/2, T/2 + 1\}$. For instance, we contaminate with a negative outlier ($\omega = -10$) at time $t = 500$ and a second one positive ($\omega = 10$) at time $t = 501$ and then we repeat the experiments by contaminating first with a positive outlier ($\omega = 10$) at time $t = 500$ and a second one negative ($\omega = -10$) at time $t = 501$. Then, we repeat this procedure with each size of the outlier considered. For each replicate, we compute the estimates of the parameters ($\alpha_0, \alpha, \beta, \delta$) by the three methods explained above and we also compute their average over all replicates. For QML and QML-t estimates, we also compute, for each replicate, the asymptotic standard deviation of each estimator and the corresponding t-ratios to test for the significance of each parameter. For comparison purposes, the LAD estimates are corrected by the corresponding scale factor $M^{-1/2}$ (see Section 2.4). Moreover, for QML-t, the parameter υ is also estimated.

3.2   Simulation results on estimation

As we are interested in the possible effect of the outliers on the leverage effect, we focus our discussion on the Monte Carlo results for the parameter δ. Actually, our results for

---

[2]The results for $T = \{500, 5000\}$, not displayed to save space, are available upon request.

the parameters α_0, α and β (not displayed here to save space) resemble those obtained in Carnero *et al.* (2007, 2012) for symmetric GARCH models. In particular, QML always overestimates α_0 and underestimates β in the presence of outliers, either isolated or in patches, and it overestimates α if there are two consecutive outliers. On the other hand, QML-t is robust to one isolated outlier without losing the good properties of QML for uncontaminated series, but it fails to be robust to patches of big outliers, especially when estimating parameters α and β. Moreover, we have also checked that the sign of the outliers does not make a great difference on estimating the parameters α_0, α and β; it is only the magnitude of the outlier rather than its sign what makes the difference. However, as Figures 2 and 3 will show, this is not the case when estimating the leverage parameter δ, where the sign of the outliers is essential.

Figure 2 displays, in its left-hand side panels, the Box-plots of the QML estimates of δ for the three models considered (one for each row) in series contaminated with one single outlier of size ω for all the sizes considered, from the most negative value ($\omega = -50$) to the highest positive one ($\omega = 50$). The central panels and the right-hand side panels of Figure 2 display the same plots for the QML-t and LAD estimates, respectively. Figure 3 displays similar Box-plots to those in Figure 2 but for series contaminated with two consecutive outliers of the same magnitude but opposite sign for all the outlier sizes considered. In particular, the labels in the x-axis of these plots represent the sign and size of the first outlier, being the 2nd one of the same magnitude but opposite sign. For instance, the Box-plot labelled as "-50" represents the distribution of the 1000 estimated values of δ from 1000 series contaminated with two consecutive outliers, namely $\{-50, 50\}$, placed at $t = \{500, 501\}$, while the Box-plot labelled as "50" corresponds to the estimated values of δ from 1000 series contaminated with two consecutive outliers, namely $\{50, -50\}$, placed at $t = \{500, 501\}$.

Several conclusions emerge from these figures. First, QML is not robust to the presence of moderate-large outliers, as expected: both the bias and dispersion of QML estimates of δ increase with the size of the outliers. Actually, in the presence of moderate-large outliers, the estimates can take any value within the admissible parameter space,

Figure 2: Boxplots of estimated δ with QML, QML-t and LAD in the presence of one isolated outlier
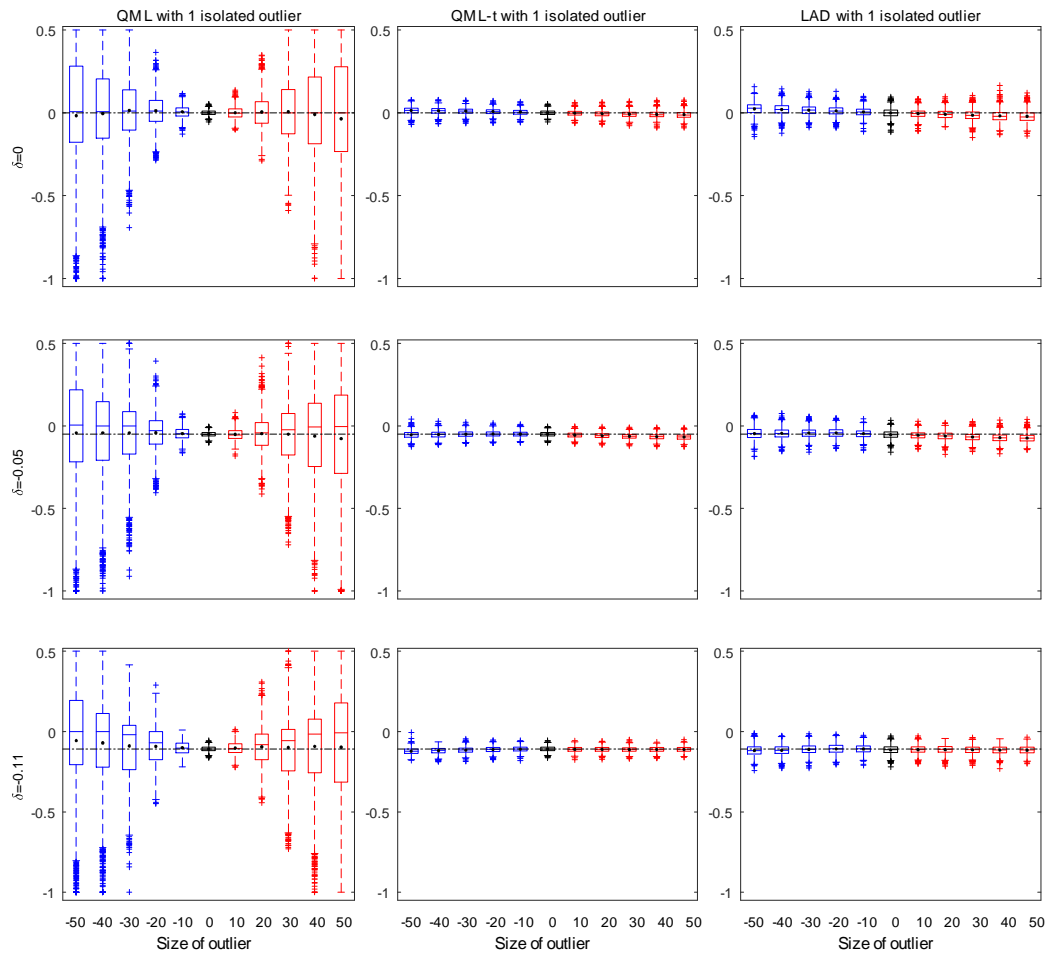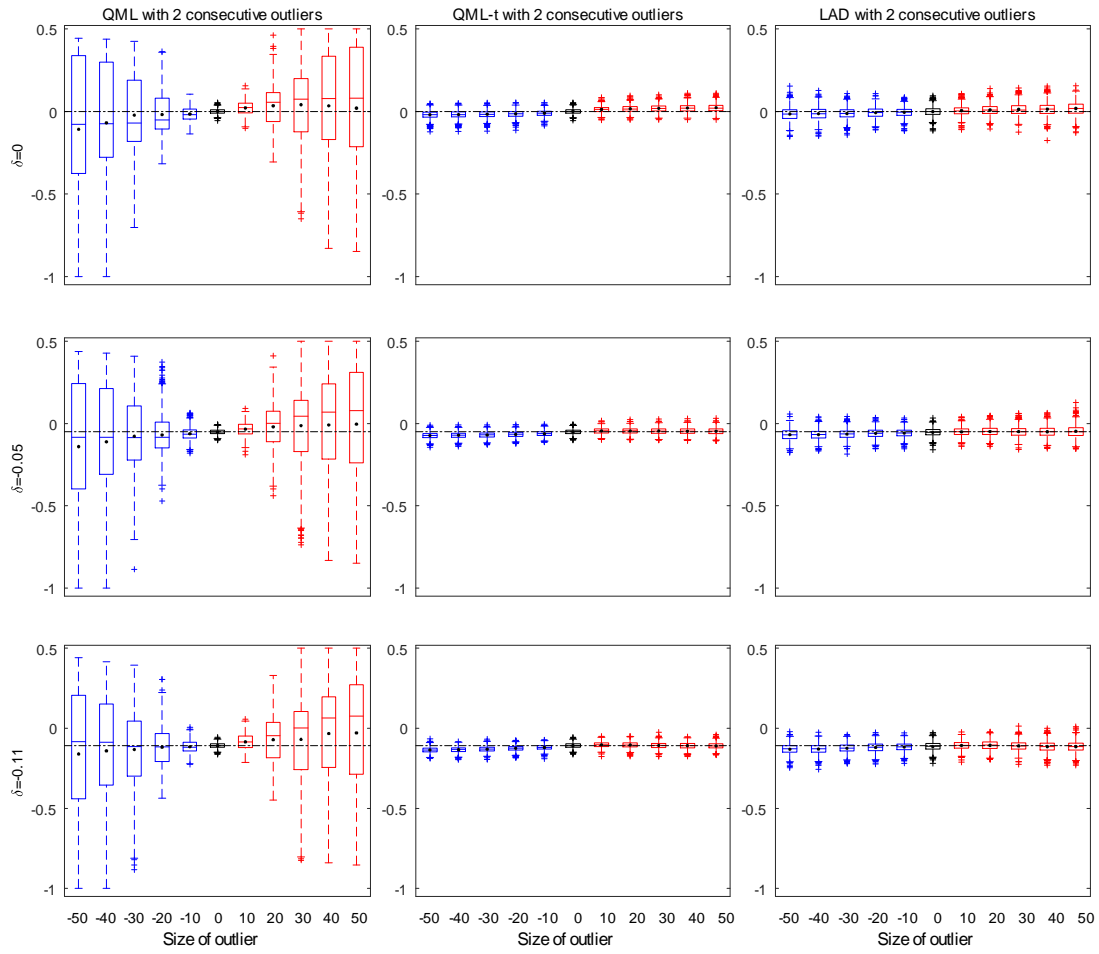
Figure 3: Boxplots of estimated δ with QML, QML-t and LAD in the presence of two consecutive outliers

rendering QML unreliable. Moreover, in models with leverage effect ($\delta \neq 0$), the presence of one isolated outlier of either sign pushes QML estimate of δ towards zero and hence could hide true leverage effect. This result agrees with the conclusions in Carnero *et al* (2016) regarding the identification of leverage based on the cross-correlogram between past returns and current squared returns. Another remarkable feature is that, in general, two consecutive outliers are more harmful in QML than one isolated outlier of the same magnitude. Comparing Figures 2 and 3, we can see that, for a given model, both the bias and the interquartile range of QML estimates are always larger in the presence of two consecutive outliers than in the presence of one isolated outlier of the same size. However, the effect of two consecutive outliers on QML estimates of δ clearly depends on the sign of the first outlier. If this is negative, QML underestimates δ and so it could yield spurious asymmetries (in models with no leverage) or it could overestimate the magnitude of the true leverage. By contrast, δ is overestimated in series contaminated with two consecutive outliers where the the 1st one is positive. Hence, in this case, QML could estimate spurious asymmetries (in models with no leverage) or it could either estimate asymmetries of the wrong sign or even no asymmetries in models with true leverage effect. Again, this agrees with the results in Carnero *et al.* (2016) regarding the impact of consecutive outliers in the identification of the leverage effect.

Figures 2 and 3 also show that both QML-t and LAD always outperform QML in the presence of outliers, as expected. Moreover, in the presence of one isolated outlier, both estimators perform quite well, even if the outlier is very large. However, they become slightly downward biased in the presence of two big consecutive outliers with the first one being negative. By contrast, when there are no outliers, QML is the best one, as expected, whereas LAD is the worst. Noticeably, QML-t does not lose much efficiency with respect to QML in this case; a similar finding is reported in Muler and Yohai (2008) for standard GARCH models.

In order to better appreciate the differences between QML-t and LAD, in Figure 4 we compare in more detail the results from these two estimators for some selected outlier sizes, namely $\omega = \{0, \pm 10, \pm 20, \pm 30\}$. As we can see, QML-t outperforms LAD in all cases, but in the presence of two consecutive outliers, it still suffers from some biases, especially if the 1st outlier is negative. Moreover, in models with no leverage or even with low leverage, the presence of two consecutive outliers, being the first one

positive, pushed QML-t estimated δ upwards leading to a possible erroneous detection of inverse leverage. Hence, there is still place to improve robustness in the estimation of asymmetric GARCH-type models, a topic that is left for further research.

3.3   Simulation results on significance tests

In this section we analyze the empirical properties of both the t-test and LR test using the standard rejection regions (11) and (12), respectively, to test $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$. We only perform this analysis with the two likelihood-based estimators, namely QML and QML-t. In all cases, the nominal size considered is 5%. In order to assess the empirical size of both tests, we simulated 1000 independent samples of size T=1000 of an AVGARCH (Parameter set 3) and, for each sample, we fitted a TGARCH model, by both QML and QML-t, and computed the corresponding t-ratio and log-likelihood. To calculate the LR statistic, we also fitted an AVGARCH and compare its log-likelihood with that of the fitted TGARCH. On the other hand, to analyze the empirical power of the tests, we simulated 1000 independent samples of size T=1000 of a TGARCH with Parameter set 2 ($\delta = -0.11$) and, for each sample, we proceed as before. We have also performed the same experiments with a TGARCH model with Parameter set 1 ($\delta = -0.05$) obtaining similar conclusions.

Figure 5 compares the empirical size of the t test (left-hand side panels) and the LR test (right-hand side panels) based on both QML and QML-t estimates of TGARCH models contaminated with 1 isolated outlier (top panels) and with two consecutive outliers (bottom panels). That is, it represents the proportion of rejections under the null ($H_0 : \delta = 0$), based on 5% critical value of t and LR tests using the standard rejection regions (11) and (12), respectively. The main conclusions we can draw from this figure are as follows. As expected, both the LR and t-test based on QML-t are always more robust to outliers than those based on QML. Moreover, as the outlier size increases, the tests based on QML become more oversized, i.e., they erroneously reject the null more often that they should and so they tend to identify spurious leverage. This problem is especially remarkable in the LR test, which could reach an over-rejection as huge as 80% in the presence of outliers of size larger than 15. By contrast, the tests

Figure 4: Boxplots of estimated δ with QML-t and LAD in the presence of one isolated and two consecutive outliers
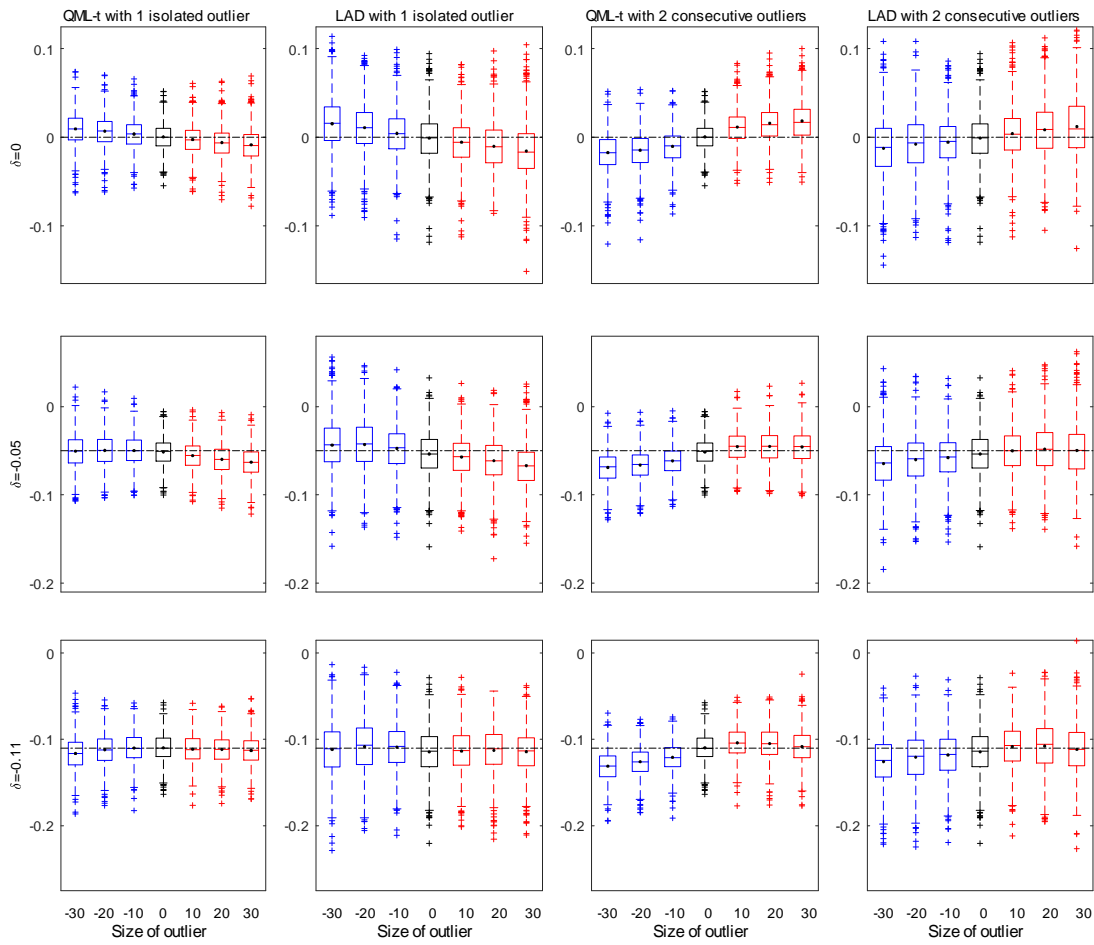
based on QML-t keep the nominal size quite well and even better in the presence of 2 outliers than in the presence on 1 isolated outlier. It is also worth mentioning that, even with no outliers, the t-test does not reach the nominal size 5%. This could be related to the warning of Francq and Zakoian (2010), mentioned in Section 2.5, regarding the inappropriateness of the standard rejection region (11) in GARCH settings. This topic is out of the scope of this paper but deserves further research.

Figure 6 compares the empirical powers of the t test (left-hand side panels) and the LR test (right-hand side panels) based on both QML and QML-t estimates of TGARCH models contaminated with 1 isolated outlier (top panels) and with two consecutive outliers (bottom panels). That is, this figure displays, for each test, the relative frequency of rejection of the hypothesis $H_0 : \delta = 0$ (no leverage) on 1000 independent realizations of length $T = 1000$ of the TGARCH model with parameter $\delta = -0.11$ (Parameter set 2). This figure shows that, in terms of power, both the LR and t-test based on QML-t are again more robust to outliers than those based on QML, as expected. The power of the tests based on QML decreases rapidly with the size of the outlier, although the loss of power is not that big in the LR test, which keeps the power around 80% when it is based on QML, even in the presence of huge outliers. However, the loss of power due to outliers of the t-test based on QML is dramatic. By contrast, the power of the tests based on QML-t is around 1 in all cases, regardless of the size, the sign and the amount of the outliers in the sample.

4   Empirical application

In this section we illustrate the previous results by fitting both the AVGARCH and the TGARCH models to two series of daily returns from different markets by using the estimation methods describe above. For each series and estimated model, we check whether the leverage parameter is significant or not by applying both the t-ratio and LR tests. The two series analyzed have been chosen to represent the possible effects than one isolated outlier and two consecutive outliers have on the inferential statistics.

Figure 5: Monte Carlo size. Proportion of rejections based on 5% critical value of t and LR tests under $H_0 : \delta = 0$
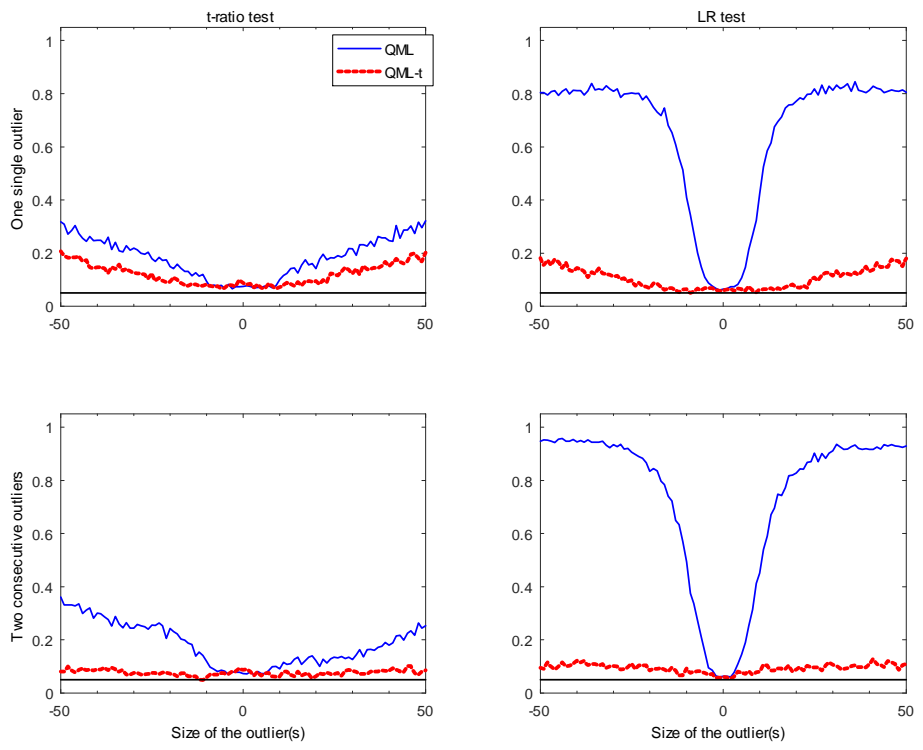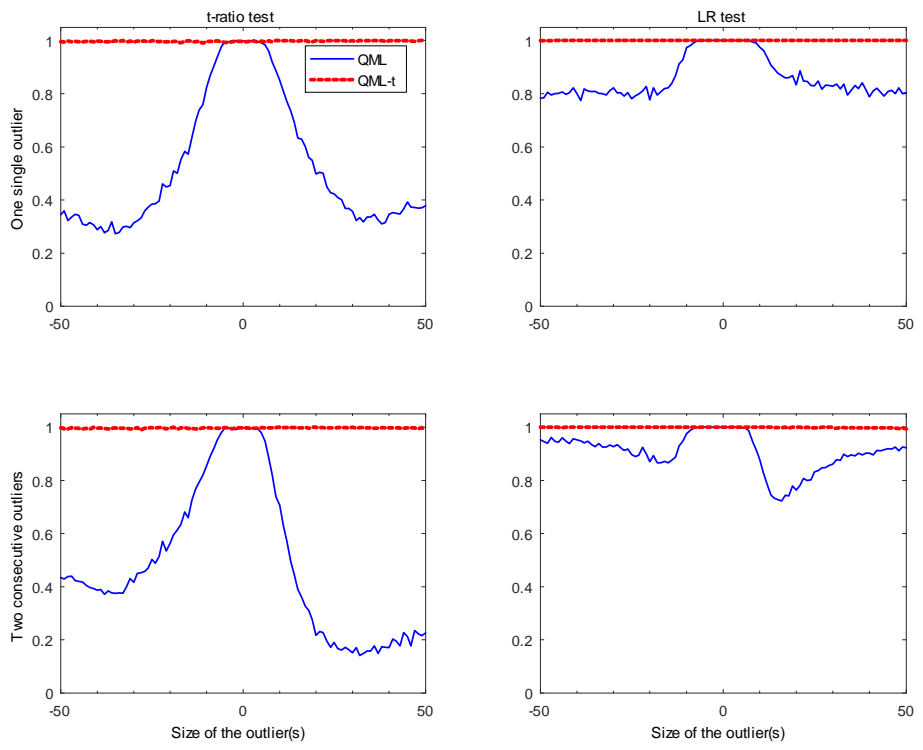
Figure 6: Monte Carlo power. Proportion of rejections based on 5% critical value of t and LR tests under $H_1 : \delta = -0.11$

4.1 Data description and dynamic properties

The series analyzed are daily close-to-close returns of the Spain IGBM Consumer Goods index observed from June 3, 2002 to December 31, 2015, comprising 3398 observations, and daily open-to-open returns of futures contracts of Natural gas from January 4, 2000 to June 28, 2013, comprising 3368 observations[3]. In both cases, the returns are computed as $y_t = 100 \times (\log(P_t) - \log(P_{t-1}))$, where P_t is the closing price at time t for the IGBM series but it represents the open price at date t for the Natural gas. The two series are plotted in Figure 7. As we can see, they both display volatility clustering and some occasional extreme values that could be regarded as outliers. For instance, the IGBM returns exhibit several extreme observations, the largest one corresponding to December 30, 2004, when the index drops by -25.3%. The Natural gas returns exhibit two extreme consecutive observations, the first one positive and the second one negative, of magnitudes about 12 times the standard deviation of the series. These outliers are due to large changes in the open price of the Natural gas on the 13th, 14th and 15th of April 2006, with values of 6.78, 11.26 and 7.15 dollars, respectively.

Table 1 contains descriptive statistics of the two series considered, as well as the Jarque-Bera and the Ljung-Box $Q(20)$ test statistics, the heteroscedastic-corrected $Q_c(20)$ test statistic proposed by Diebold (1988), and the test statistic $CH(20)$, proposed by Cumby and Huizinga (1992) which is also robust to conditional heteroscedasticity. We also include the values of the statistics Q and CH applied to the squared returns, denoted by $Q_2(20)$ and $CH_2(20)$, respectively, to test for conditional heteroscedasticity. As expected, both series exhibit excess kurtosis and the Jarque-Bera test for Normality always rejects the null. The values of $Q_c(20)$ and $CH(20)$ never reject the null at 5% significance level, suggesting that both returns are uncorrelated, as expected. Surprisingly, the values of $Q_2(20)$ and $CH_2(20)$ are not significant for the IGBM series, suggesting that its conditional variance is constant over time, while the values of these test statistics are significant at 1% for the Natural gas, indicating strong evidence of conditional heteroscedasticity.

---

[3]The data for the IGBM Consumer Goods index was downloaded from www.invertia.com and the corresponding one for Natural gas was obtained from the additional files in Kristoufek (2014).

Figure 7: Daily returns for IGBM Consumer Goods index and Natural gas
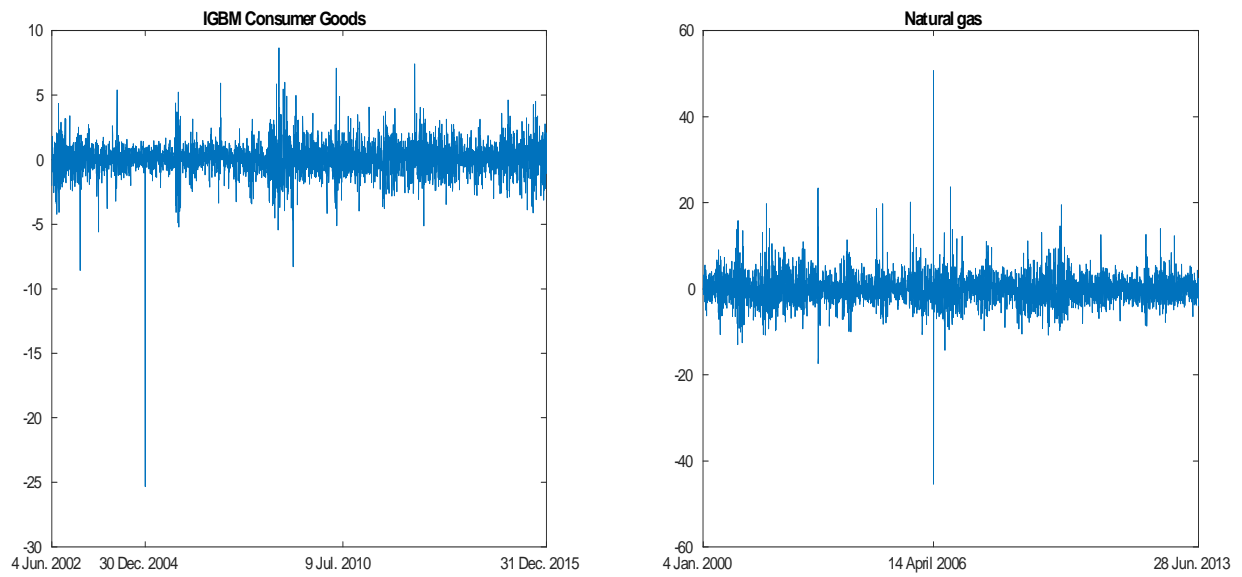
Table 1: Descriptive statistics of the returns

Statistic	IGBM Consumer Goods	Natural gas
Mean	0.0439	0.0149
Std.Dev.	1.3320	3.7653
Skewness	-2.0142	0.7462
Kurtosis	44.542	21.765
Jarque-Bera	246630***	4973***
Sample size	3398	3368
$Q(20)$	37.537**	48.547***
$Q_c(20)$	28.989*	18.593
$CH(20)$	28.956*	18.975
$Q_2(20)$	2.9310	631.56***
$CH_2(20)$	23.119	75.94***

*,**,***: statistically significant at 10%, 5% and 1% respectively

Figure 8 displays, in its first row, the correlograms of the returns with the corrected 95% confidence bands proposed by Diebold (1988) for conditionally hereroscedastic series. These bands are wider than the usual Barlett bands and show no evidence of autocorrelated returns. The correlogram of squared returns, displayed in the 2nd row of Figure 8, suggests no correlation in the squared returns of IGBM, looking like the typical correlogram of a white noise series and confirming the previous result on the tests $Q_2(20)$ and $CH(20)$. Nevertheless, it should be noticed that the pattern of this correlogram can also be due to the harmful effect of one isolated outlier on conditionally heteroscedastic series; see Carnero *et al.* (2007). On the other hand, for the Natural gas, we can see the typical pattern of the correlogram of the squared observations in the presence of two consecutive outliers, that is, a very high positive and significant 1st order correlation while the others are pushed downwards towards zero; see Carnero *et al.* (2007). However, when the robust autocorrelations of squares proposed by Teräsvirta and Zhao (2011) are computed (see the 3rd row of Figure 8), another picture comes up. For both, the IGBM and the Natural gas, all these correlations become significantly different from zero, suggesting that the conditional variances of both returns are not constant over time, as expected, and confirming that the patterns of the sample autocorrelations are due to the presence of outliers.

Finally, the last two rows of Figure 8 display the sample cross-correlations between past and squared returns and their robust counterparts, respectively. The latter are computed using the proposal in Carnero *et al.* (2016) based on applying the Ramsay weighting scheme to the sample variances and cross-covariances. Again, the picture changes depending on whether we look at the sample or the robust cross-correlations. For the IGBM returns, the sample cross-correlations suggest no leverage effect. By contrast, most of the robust cross-correlations are negative suggesting possible leverage effect in this series. As discussed in Carnero *et al.* (2016), this is the expected pattern due to one isolated outlier, which biases all sample cross-correlations towards zero hiding true leverage. On the other hand, for the Natural gas, the patterns are quite different,

Figure 8: Correlograms of returns and squared returns and cross-correlograms between lagged returns and squared returns
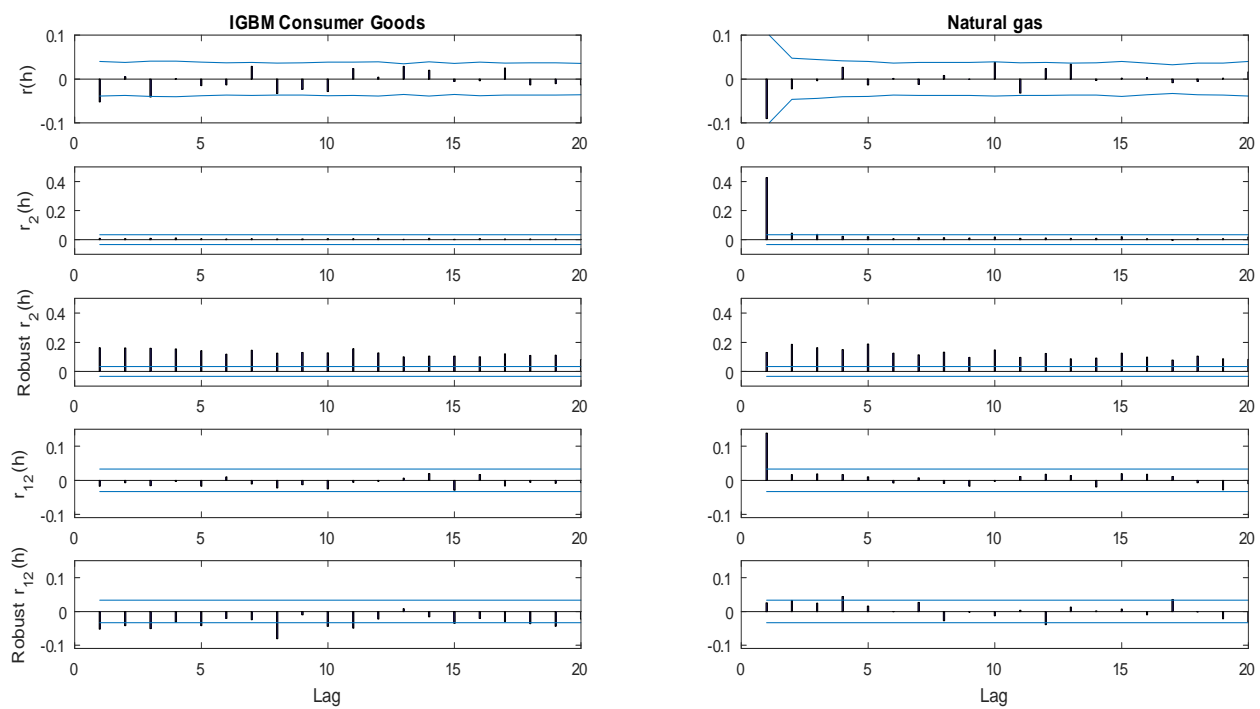
whereas all the robust cross-correlations are around zero, indicating no leverage effect, the 1st sample cross-correlation is pushed upwards to a significant positive value, suggesting inverse leverage effect (positive relationship between volatility and past returns). Again, this is the expected pattern due to the presence of two consecutive outliers, being the first one positive; see Carnero *et al.* (2016). Therefore, the inverse leverage effect found in the Natural gas by some authors (see Kristoufek (2014) and the references therein) could be an artifact due to the misleading effect of outliers.

4.2 Estimating and testing for leverage effect

We describe below the results from estimating the TGARCH and AVGARCH models to the two return series described above. Since the results in Section 3 suggest a better performance of QML-t over LAD and the computation of the standard errors of the latter is cumbersome, we focus our comparison on QML and QML-t, although the estimated leverage parameter by LAD will be also commented. Table 2 displays the estimation results obtained by QML and QML-t as well as some diagnostics based on the residuals, $\widehat{\varepsilon}_t = y_t/\widehat{\sigma}_t$, where $\widehat{\sigma}_t$ is the estimated volatility for each model. The estimation has been carried out with the Oxford MFE Toolbox for Matlab, taking into account the reparametrization used in this package (see Appendix A) in order to compute the estimated parameter values and their corresponding standard errors.[4]

As we can see, there are remarkable differences between the results obtained for the two returns series and, for each series, there are also some differences between both estimation methods that are worth mentioning. Since we are interested in the effects of outliers on estimating and testing the leverage effect, our comments will be focused on the leverage parameter δ. For the IGBM series, we can see that, regardless of the estimation method, the parameter δ in the TGARCH model is estimated negative, as it is expected if there is leverage effect. The LAD estimate is also negative $\left(\widehat{\delta}_{LAD} = -0.0229\right)$. However, the leverage coefficient is not statistically significant when the

---

[4]To check for the robustness of our results, we have repeated the estimation by QML and QML-t in Stata, obtaining similar results. We have also estimated EGARCH models leading to similar conclusions.

Table 2: Estimation of the TGARCH and AVGARCH models with QML and QML-t

Estimator	Parameter	IGBM Consumer Goods		Natural gas	
		TGARCH	AVGARCH	TGARCH	AVGARCH
QML	ω	0.1041** (0.053)	0.0919** (0.043)	0.1019* (0.059)	0.1063 (0.071)
	α	0.0500* (0.030)	0.0515* (0.031)	0.0749** (0.029)	0.0794** (0.036)
	β	0.8858*** (0.018)	0.8942*** (0.012)	0.9172*** (0.035)	0.9128*** (0.043)
	δ	−0.0171 (0.013)		0.0077 (0.012)	
	Log-Likelihood	−5697.7	−5700.8	−9028.7	−9029.9
Residuals	$Q(20)$	25.515	24.981	22.239	22.167
	$CH(20)$	29.660*	29.134*	15.991	15.875
	$Q_2(20)$	0.138	0.140	51.178***	57.708***
	$CH_2(20)$	2.910	2.835	11.688	12.460
QML-t	ω	0.0292*** (0.009)	0.0310*** (0.011)	0.0833*** (0.023)	0.0788*** (0.023)
	α	0.0911*** (0.014)	0.0982*** (0.015)	0.0806*** (0.010)	0.0813*** (0.010)
	β	0.9087*** (0.016)	0.9016*** (0.018)	0.9144*** (0.012)	0.9152*** (0.012)
	δ	−0.0373*** (0.009)		0.0185** (0.007)	
	υ	5.2918	5.2311	6.7494	6.7402
	Log-Likelihood	−5191.7	−5203.9	−8779.8	−8784.0
Residuals	$Q(20)$	14.520	15.451	20.864	21.393
	$CH(20)$	27.645	27.182	16.273	15.904
	$Q_2(20)$	0.113	0.151	25.894	47.320**
	$CH_2(20)$	7.642	7.967	12.371	14.130
*,**,***: statistically significant at 10%, 5%, and 1%, respectively					

QML-based t-ratio test is applied but becomes significant when QML-t is used. In particular, the values (and p-values) of the t-statistic are -1.330 (0.1836) and -4.276 (0.000), respectively. Meanwhile, the values (and p-values) of the LR test statistic for $H_0 : \delta = 0$ (AVGARCH) against $H_1 : \delta \neq 0$ (TGARCH) are 6.248 (0.0124) and 24.356 (0.000) when the models are estimated by QML and QML-t, respectively, confirming that δ is significant when QML-t is used but it is not longer significant at 1% when QML is applied. This result agrees with our discussion in Section 3 where we show that a single isolated outlier (like the one existing in the IGBM series) biases the QML estimated δ towards zero and could hide true leverage. Then, it seems that, in this case, QML-t is more reliable than QML suggesting that there is leverage effect in the IGBM returns, although this is hidden by the huge return observed on December 30, 2004, when the model is estimated by QML.

For the Natural gas series, the parameter δ is estimated positive by both QML and QML-t and also by LAD ($\widehat{\delta}_{LAD} = 0.0217$), suggesting inverse leverage effect. However, its statistical significance also depends on the estimator used as well as the test statistic chosen. For this series, the values (and p-values) of both the t-ratio and LR test statistics are 0.636 (0.5247) and 2.377 (0.123), respectively, when QML is used, indicating no evidence of leverage effect at any reasonable significance level. But, when the model is estimated using QML-t, the values (and p-values) of these two test statistics are 2.550 (0.0108) and 8.458 (0.0036), respectively, showing evidence of inverse leverage effect at 5% significance level (or even at 1% if the LR test is selected). This somehow surprising result could be related to our findings in Section 3, where we show that QML-t could be still biased in the presence two big consecutive outliers, with the estimated δ being pushed upwards if the first of these outliers is positive, as it is the case in the Natural gas series. Therefore, we should be very careful in concluding that there is actually inverse leverage effect in this series, as this could be a misleading effect caused by the two consecutive extreme observations present in this series. Finally, it is also worth mentioning that the thickness parameter υ of the Student error distribution in QML-t, indicates fat tails in both returns.

When looking at residuals diagnostics, as expected, the values of the test statistics $Q_2(20)$ and $CH_2(20)$ for remaining autocorrelation in the squared residuals, have been reduced remarkably in all estimated models, as compared to their values for the

returns displayed in Table 1. Only for the Natural gas, the values of $Q_2(20)$ remain significant at 1% when the two models are estimated by QML. However, when QML-t is used, both statistics $Q_2(20)$ and $CH_2(20)$ are not longer significant for the TGARCH model, suggesting that this model has been able to properly capture the dynamics in the conditional variance of the returns.

To further illustrate how the potential outliers can bias the estimation and testing of the leverage effect, we consider a rolling window scheme of size $T^* = 1000$ through the whole sample, starting at point $t = 1$ and moving forwards up to covering the last 1000 observations of the full sample. For the IGBM and the Natural gas series, this amounts to analyzing 2398 and 2368 subsamples, respectively, covering periods of different volatilities and types and sizes of outliers, according to the following steps:

1. Select the 1st subsample for IGBM (04/06/2002 - 31/07/2006) and for Natural gas (04/01/2000 - 9/01/2004), estimate the parameter δ by the three methods considered (QML, QML-t and LAD) and compute, for QML and QML-t, the corresponding t-ratio and LR test statistics for these subseries.

2. Delete the 1st observation, add a new observation at the end of the subsample and re-estimate the model and test again for leverage.

3. Repeat the process until we reach the end of the full samples, where we cover the last subsamples for IGBM (26/01/2012 - 31/12/2015) and Natural gas (23/06/2009 - 28/06/2013).

Figure 9 plots, in its first row, the estimated values of δ across the rolling window for the two series considered and for the three estimation methods applied, i.e., it plots the values of $\widehat{\delta}_{QML}$, $\widehat{\delta}_{QML-t}$ and $\widehat{\delta}_{LAD}$ for each of the 2398 subseries of the IGBM series (left-hand side panel) and for each of the 2368 subseries of the natural gas series (right-hand side panel). As a benchmark, the zero line is also displayed to account for no leverage. Similarly, Figure 9 plots the corresponding p-values of both the t-ratio (2nd row) and the LR (3rd row) tests to test $H_0 : \delta = 0$ against $H_0 : \delta \neq 0$. The horizontal line represents the 5% significance level and consequently, those subsamples with p-values below this line are those where the leverage effect is significant at 5% significance level.

Figure 9: Leverage parameter δ in a TGARCH model estimated with subsamples of size T = 1000 using a rolling window of both the IGBM Consumer Goods index and Natural gas daily returns (top panels) and the corresponging p-values of the t-ratio and LR test statistics to test for the statistical significance of δ.
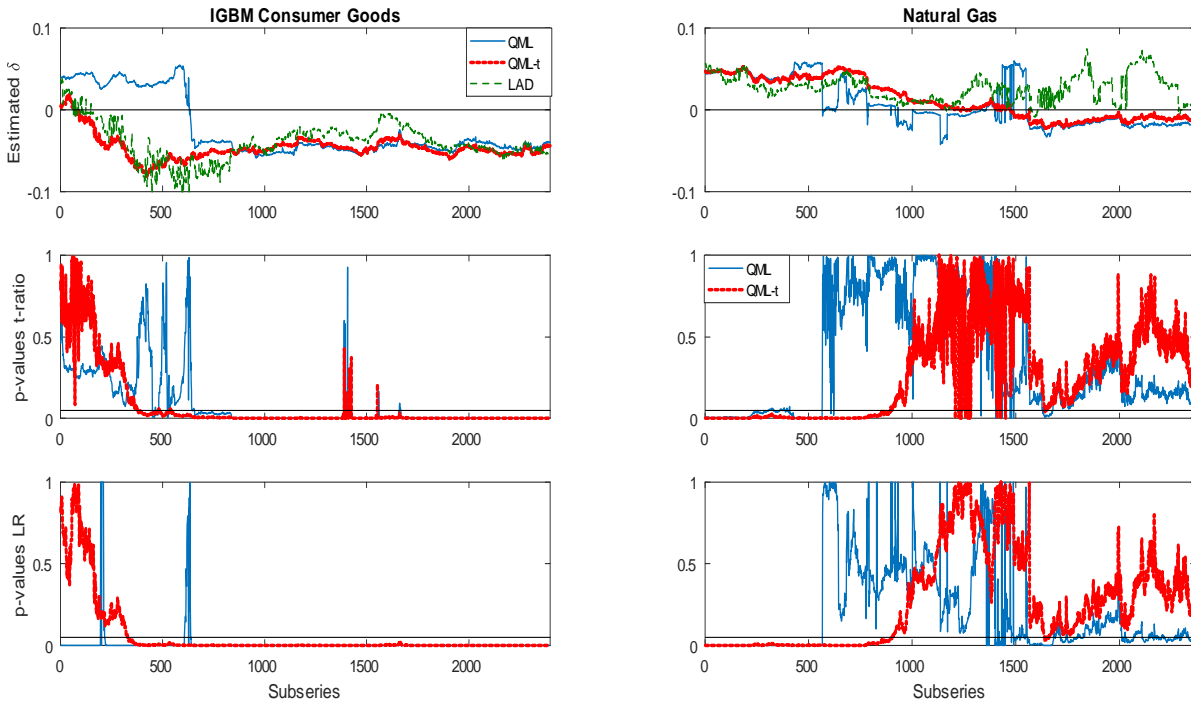
Several conclusions emerge from this figure. First, we observe how extreme observations can bias the estimated leverage parameter and the corresponding test statistics and could lead to a wrong conclusion about the sign and magnitude of the leverage effect. As expected, the QML estimated values of δ present several sharp drops and rises which do not appear in their robust counterparts. These sharp changes are usually due to the entrance and/or an exit of outlying observations in the corresponding subsample. For instance, in the IGBM series, the exit of the observation in December 30, 2004, where the index sustained its largest drop ($y_{642} = -25.33$), conveys a sudden fall in $\widehat{\delta}_{QML}$ from a positive value to a negative value around -0.035 (see the graph in the top left-hand side panel). Noticeably, for the first subseries corresponding at the beginning of the full sample, i.e., for those including that isolated outlier, $\widehat{\delta}_{QML}$ is always positive but not significant at any reasonable significance level, as judged by the t-ratio test (see the corresponding p-values displayed in the left panel in the 2nd row of Figure 9). Moreover, for these subsamples, there are big differences between $\widehat{\delta}_{QML}$ and the robust estimates, $\widehat{\delta}_{QML-t}$ and $\widehat{\delta}_{LAD}$, which are both both mainly negative and very similar to each other. Actually, when QML-t is used, the p-values of both significance tests show that δ is statistically significant for several subsamples, indicating the presence of leverage effect. Recall that, according to our results in Section 3, these differences are expected in the presence of one isolated outlier, which seems to be the cause here of the upwards bias observed in $\widehat{\delta}_{QML}$, as compared to the robust estimators.

It is also worth noting that, when we look at the patterns after the exit of that outlying negative observation of the IGBM index, i.e. when the subseries analyzed do not longer contain this outlier, the differences between the three estimators (QML, QML-t and LAD) are considerably reduced, specially for the last subseries at the end of the sample. Actually, for these subseries, which are free from outliers (see the calm period in the IGBM returns displayed in Figure 7 at the end of the sample), we observe that, regardless of the estimator and the test statistics used, δ is always estimated negative and statistically significant indicating the presence of leverage effect. Again, this result agrees with our findings in Section 3 for the simulated series without outliers.

When looking at the results for the Natural gas (right-hand side panels), we observe remarkable differences among the three estimators considered, specially for the subseries located around the middle of the sample, i.e., those including the two huge consecutive

outliers present in this series ($y_{1566} = 50.73$ and $y_{1567} = -45.41$). In these cases, our results in Section 3 show that even the robust estimators could still be slightly biased. In particular, when these two observations, the first one being positive and the next one negative, are in a subsample, we expect $\widehat{\delta}_{QML}$ but also $\widehat{\delta}_{QML-t}$ to be upwards biased. Another remarkable feature from the Natural gas results in Figure 9, is that the values of $\widehat{\delta}_{QML}$ and $\widehat{\delta}_{QML-t}$ for the subseries at the end of the period (see the graph in the top right-hand side panel), are very similar to each other but quite different from the values of $\widehat{\delta}_{LAD}$: the former take negative values (standard leverage) whereas the latter estimate $\delta > 0$ (inverse leverage). This feature could be related to the fact that LAD is underperforming if no outliers are present in the series, as discussed in Section 3. Accordingly, for most of the subseries at the end of the period, neither the t-ratio nor the LR test statistics based on QML and QML-t reject the null, $H_0 : \delta = 0$, in these subseries. Therefore, we wonder whether the inverse leverage effect found in the Natural gas returns by some authors (see, for example, Kristoufek (2014) and the references therein) could be due to the harmful effect of these consecutive outliers.

Alternatively, the patterns of the estimated parameters displayed in Figure 9 could be indicating that the leverage effect is time-varying, as suggested by Bandi and Reno (2012), Yu (2012) and Jensen and Maheu (2014), but this is a different problem that is out of the scope of this paper.

5   Conclusions

This paper has analyzed the effects of outliers on the estimation and testing for the leverage effect in the context of the TGARCH model. It is shown that, as expected, QML-t and LAD always outperform QML in the presence of outliers. Actually, one isolated outlier could lead QML to hide true leverage effect whereas two consecutive outliers bias the QML estimated leverage in a direction that crucially depends on the sign of the first outlier. If this is negative (positive), QML underestimates (overestimates) the leverage parameter. Therefore, in these cases, QML could hide true leverage or estimate spurious asymmetries or asymmetries of the wrong sign. However, both robust estimators perform very well when there is one isolated outlier, but they are still slightly biased in the presence of patches of big outliers, leading, in some cases, to inaccurate

estimates of the leverage coefficient. In general, QML-t seems to outperform LAD because it is robust to moderate-big outliers without losing much efficiency, as compared to QML, when there are no outliers. That is not the case for LAD, which performs much worse than QML with no outliers. These results are further illustrated with the empirical analysis of two return series from financial and energy markets, including one isolated negative outlier and two consecutive outliers, respectively.

Appendix A. Parametrizations of the TGARCH(1,1) model

There are several (equivalent) parametrizations of the TGARCH(1,1) model proposed in the econometric literature and also in the econometric software packages. Hence, in order to compare estimates and standard errors from different papers and/or different software packages, it is important to be very careful about which parametrization is being used in each case. The one we have adopted in this paper is taken from Rodríguez and Ruiz (2012) and it is given by the following equations:

$$y_t = \sigma_t \, \varepsilon_t,$$

$$\sigma_t = \alpha_0 + \alpha \, |y_{t-1}| \, + \beta \sigma_{t-1} + \delta y_{t-1}. \tag{A.1}$$

Hence, the volatility equation (A.1) can be re-written as:

$$\sigma_t = \begin{cases} \alpha_0 + (\alpha + \delta) \, y_{t-1} \, + \beta \sigma_{t-1} & \text{if} \quad y_{t-1} \geq 0 \\ \alpha_0 - (\alpha - \delta) \, y_{t-1} \, + \beta \sigma_{t-1} & \text{if} \quad y_{t-1} < 0 \end{cases}. \tag{A.2}$$

In turn, the original proposal in Zakoian (1994) defines the volatility as follows:

$$\sigma_t = \alpha_0 + \alpha^+ y_{t-1}^+ \, - \alpha^- y_{t-1}^- \, + \beta \sigma_{t-1}, \tag{A.3}$$

where:

$$y_t^+ = \max(y_t, 0), \quad y_t^- = \min(y_t, 0).$$

Hence, the volatility equation (A.3) can be re-written as:

$$\sigma_t = \begin{cases} \alpha_0 + \alpha^+ y_{t-1} \, + \beta \sigma_{t-1} & \text{if} \quad y_{t-1} \geq 0 \\ \alpha_0 - \alpha^- y_{t-1} \, + \beta \sigma_{t-1} & \text{if} \quad y_{t-1} < 0 \end{cases}. \tag{A.4}$$

Therefore, comparing (A.2) and (A.4), the following relationship comes up between the parameters of both parametrizations:

$$\begin{matrix} \alpha + \delta = \alpha^+ \\ \alpha - \delta = \alpha^- \end{matrix} \iff \begin{matrix} \alpha = \frac{1}{2}(\alpha^+ + \alpha^-) \\ \delta = \frac{1}{2}(\alpha^+ - \alpha^-) \end{matrix}.$$

He and Teräsvirta (1999) consider the same volatility specification as Zakoian (1994), but they write it down in a slightly different way, namely

$$\sigma_t = \alpha_0 + \sigma_{t-1}\{\beta + [\alpha^+(1 - I(\varepsilon_{t-1})) + \alpha^- I(\varepsilon_{t-1})] \, |\varepsilon_{t-1}|\}, \tag{A.5}$$

where

$$I(\varepsilon_{t-1}) = \begin{cases} 1 & \text{if} \quad \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}.$$

Hence, taking into account that $y_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$ and $sign(y_{t-1}) = sign(\varepsilon_{t-1})$, the volatility equation (A.5) can be easily re-written as in (A.4).

Another proposal, very similar to that in Rodríguez and Ruiz (2012), can be found in He *et al.* (2008), whose volatility equation is as follows

$$\sigma_t = \alpha_0 + \sigma_{t-1}\{\beta + \alpha \ |\varepsilon_{t-1}| + \alpha^*\varepsilon_{t-1}\}. \tag{A.6}$$

Hence, taking into account that $y_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$, the volatility equation (A.6) can be re-written as in (A.1) with $\alpha^* = \delta$.

Finally, the MFE Toolbox for Matlab that we have used in our Monte Carlo experiments and empirical application, employs the following volatility equation for the TGARCH(1,1) model:

$$\sigma_t = \alpha_0 + \alpha_1|y_{t-1}| \ + \gamma_1|y_{t-1}| \ I(y_{t-1} < 0) + \beta\sigma_{t-1}, \tag{A.7}$$

where

$$I(y_{t-1} < 0) = \begin{cases} 1 & \text{if} \quad y_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}.$$

Hence, the volatility equation (A.7) can be re-written as:

$$\sigma_t = \begin{cases} \alpha_0 + \alpha_1 \ y_{t-1} + \beta\sigma_{t-1} & \text{if} \quad y_{t-1} \geq 0 \\ \alpha_0 - (\alpha_1 + \gamma_1) \ y_{t-1} \ + \beta\sigma_{t-1} & \text{if} \quad y_{t-1} < 0 \end{cases}. \tag{A.8}$$

Therefore, comparing (A.2) and (A.8), the following relationships come up:

$$\begin{matrix} \alpha + \delta = \alpha_1 \\ \alpha - \delta = \alpha_1 + \gamma_1 \end{matrix} \quad \Longleftrightarrow \quad \begin{matrix} \alpha = \alpha_1 + \frac{1}{2}\gamma_1 \\ \delta = -\frac{1}{2}\gamma_1 \end{matrix}.$$

These relationships should be taken into account to properly compute the correct point estimates and standard errors of our parameters $(\alpha_0, \alpha, \beta, \delta)'$ from the estimated parameters computed by the MFE Toolbox, namely $(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\beta}, \widehat{\gamma}_1)$, and their asymptotic variance-covariance matrix.

Another parametrizations of the TGARCH(1,1) model are still possible; see, for instance, Hentschel (1995).

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001) The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43–76.

[2] Ait-Sahalia, Y., Fan, J. and Li, Y. (2013) The leverage effect puzzle: Disentangling sources of bias at high frequency. *Journal of Financial Economics* 109, 224–249.

[3] Bandi, F. and Reno, R. (2012) Time-varying leverage effects. *Journal of Econometrics* 169, 94–113.

[4] Berkes, I. and Horvath, L. (2004) The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics* 32, 633–655.

[5] Black, R. (1976) Studies in stock price volatility changes, Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Sections, American Statistical Association, 177–181.

[6] Carnero, M.A., Peña, D. and Ruiz, E. (2007) Effects of outliers on the identification and estimation of GARCH models. *Journal of Time Series Analysis* 28, 471–497.

[7] Carnero, M.A., Peña, D. and Ruiz, E. (2012) Estimating GARCH volatility in the presence of outliers. *Economics Letters* 114, 86–90.

[8] Carnero, M.A., Pérez, A. and Ruiz, E. (2016) Identification of asymmetric conditional heteroscedasticity in the presence of outliers. *SERIEs: Journal of the Spanish Economic Association* 7, 179 –201.

[9] Chkili, W., Hammoudeh, S. and Nguyen, D. (2014) Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. *Energy Economics* 41, 1–18.

[10] Creal, D., Koopman, S. J. and Lucas, A. (2013) Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28, 777-795.

[11] Cumby, R. and Huizinga, J. (1992) Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* 60, 185–195.

[12] Diebold, F.X. (1988) *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag.

[13] Ding, Z., Granger, C. and Engle, R. (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

[14] Engle, R.F. (2011) Long term skewness and systematic risk. *Journal of Financial Econometrics* 9, 437–468.

[15] Fan, J., Qi, L. and Xiu, D. (2014) Quasi-Maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics* 32, 178–191.

[16] Francq, C. and J-M. Zakoian (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

[17] Francq, C. and J-M. Zakoian (2013) Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statististical Society B* 75, 345–367.

[18] Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48, 1779–1801.

[19] Harvey, A. C. (2013) *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press.

[20] He, C. and Teräsvirta, T. (1999) Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.

[21] He, C., Silvennoinen, A. and Teräsvirta T. (2008) Parameterizing unconditional skewness in models for financial time series. *Journal of Financial Econometrics* 6, 208–230.

[22] Hentschel, L. (1995) All in the family Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39, 71–104.

[23] Hibbert, A.M., Daigler, R.T. and Dupoyet, B. (2008) A behavioral explanation for the negative asymmetric return-volatility relation. *Journal of Banking and Finance* 32, 2254–2266.

[24] Hill, J.B. (2015) Robust estimation and inference for heavy tailed GARCH. *Bernoulli* 21, 1629–1669.

[25] Huang, D., Wang, H. and Yao, Q. (2008) Estimating GARCH models: when to use what? *Econometrics Journal* 11, 27–38.

[26] Hwang, S. and Basawa, I. (2004) Stationarity and moment structure for Box-Cox transformed threshold GARCH(1,1) processes. *Statistics & Probability Letters* 68, 209–220.

[27] Jensen, M.J. and Maheu, J.M. (2014) Estimating a semiparametric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics* 178, 523–538.

[28] Kristoufek, L. (2014) Leverage effect in energy futures. *Energy Economics* 45, 1–9.

[29] Laurent, S., Lecourt, C. and Palm, F.C. (2016) Testing for jumps in conditionally Gaussian ARMA-GARCH models, a robust approach. *Computational Statistics & Data Analysis* 100, 383–400.

[30] Mendes, B. (2000) Assessing the bias of maximum likelihood estimates of contaminated garch models. *Journal of Statistical Computation and Simulation* 67, 359–376.

[31] Muler, N. and Yohai, V. (2008) Robust estimates for GARCH models. *Journal of Statistical Planning and Inference* 138, 2918–2940.

[32] Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.

[33] Newey, W. K. and Steigerwald, D.G. (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity. *Econometrica* 65, 587–599.

[34] Pan, J., Wang H. and Tong H. (2008) Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* 142, 352–378.

[35] Peng, L. and Yao, Q. (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* 90, 967–975.

[36] Rodríguez, M.J. and Ruiz E. (2012) GARCH models with leverage effect: differences and similarities. *Journal of Financial Econometrics* 10, 637–668.

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

[38] Schwert, G.W. (1989) Why does stock market volatility change over time? *Journal of Finance* 45, 1129–1155.

[39] Sentana, E. (1995) Quadratic ARCH Models. *Review of Economic Studies* 62, 639–661.

[40] Straumann, D. and Mikosch T. (2006) Quasi-maximum-likelihood estimation in heteroskedastic time series: A stochastic recurrence equations approach. *Annals of Statistics* 34, 2449–2495.

[41] Taylor, S.J. (1986) *Modelling Financial Time Series*. Wiley.

[42] Teräsvirta, T. and Zhao Z (2011) Stylized facts of return series, robust estimates and three popular models of volatility. *Applied Financial Economics* 21, 67–94.

[43] Wintenberger, O. (2013) Continuous invertibility and stable QML estimation of the EGARCH(1,1) model. *Scandinavian Journal of Statistics* 40, 846–867.

[44] Yu, J. (2012) A semiparametric stochastic volatility model. *Journal of Econometrics* 167, 473–482.

[45] Zakoian, J.M. (1994) Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

[46] Zivot, E. (2009) Practical issues in the analysis of univariate GARCH models. In: Mikosch T, Kreiß J-P, Davis RA, Andersen TG (eds) *Handbook of Financial Time Series.* Springer-Verlag, Berlin, pp 113–155.

serie
ad