Robust estimation with the weighted trimmed likelihood estimator

Chalabi, Yohan and Wuertz, Diethelm

Institute of Theoretical Physics, ETH Zurich, Switzerland, Computational Science and Engineering, ETH Zurich, Switzerland

November 2012

Weighted Trimmed Likelihood Estimator for GARCH Models

Yohan Chalabi[a,b,*], Diethelm Würtz[a,b]

[a] *Institute of Theoretical Physics, ETH Zurich, Switzerland*
[b] *Computational Science and Engineering, ETH Zurich, Switzerland*

Abstract

Generalized autoregressive heteroskedasticity (GARCH) models are widely used to reproduce stylized facts of financial time series and today play an essential role in risk management and volatility forecasting. But despite extensive research, problems are still encountered during parameter estimation in the presence of outliers. Here we show how this limitation can be overcome by applying the robust weighted trimmed likelihood estimator (WTLE) to the standard GARCH model. We suggest a fast implementation and explain how the additional robust parameter can be automatically estimated. We compare our approach with other recently introduced robust GARCH estimators and show through the results of an extensive simulation study that the proposed estimator provides robust and reliable estimates with a small computation cost. Moreover, the proposed fully automatic method for selecting the trimming parameter obviates the tedious fine tuning process required by other models to obtain a "robust" parameter, which may be appreciated by practitioners.

Keywords: GARCH Models, Robust Estimators, Outliers, Weighted Trimmed Likelihood Estimator (WTLE), Quasi Maximum Likelihood Estimator (QMLE)

*Corresponding address: HIT K 21.2, Wolfgang-Pauli-Str. 27, 8093 Zurich. Tel.: +41 44 633 70 53

Email address: chalabi@phys.ethz.ch (Yohan Chalabi)

1. Introduction

Because time-variation in the volatility is a characteristic feature of financial time series, accurate modeling of this feature is critical in many financial applications and especially so in risk management. Since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982) and its generalization, known as the GARCH model by Bollerslev (1986), there has been a large amount of theoretical and applied research work concerning these models. The success of GARCH models stems mainly from their ability to reproduce the typical stylized facts of financial time series, particularly, volatility clustering, the fat-tail distribution of financial returns, and the long-term memory effect. Additionally, GARCH processes can be modeled with a wide range of innovation distributions and can be tailored to specific problems. Indeed, Bollerslev (2008) compiled a glossary of more than 150 GARCH models. GARCH modeling is now common in practice, but this is despite the fact estimation of its parameters involves solving a rather difficult constrained nonlinear optimization problem. Moreover, it is common for different software implementations to leave one with a set of conflicting estimates (Brooks et al., 2001). And besides the difficulty in parameter estimation, GARCH models remain, as any other model, approximations that cannot be expected to encompass all of the complex dynamics of financial markets: Market conditions are strongly affected by factors such as rumor, news, speculation, policy changes, and even data recording errors, which can result in abnormal points, or outliers, that are beyond the scope of the model. Moreover, the maximum likelihood estimator (MLE) of GARCH models is very sensitive to outliers (Mendes, 2000; Hotta and Tsay, 1998).

The weighted trimmed likelihood estimator (WTLE), introduced by Hadi and Luceño (1997) and Vandev and Neykov (1998), is a generalization of the trimmed likelihood estimator (TLE) of Bednarski and Clarke (1993). Since its introduction, the WTLE has been applied in many different fields: Markatou (2000) used the WTLE for mixture models, Müller and Neykov (2003) studied related estimators in generalized linear models, and Neykov et al. (2007) employed the WTLE for robust parameter estimation in a finite mixture of distributions. Bednarski and Clarke (1993) discuss the Fisher consistency, compact differentiability, and asymptotic normality of the TLE. Cizek (2008) explores the consistency and asymptotic properties of the WTLE. Also, see (Vandev and Neykov, 1998; Müller and Neykov, 2003; Dimova and Neykov, 2004) for the derivation of the breakdown point of the WTLE for various models.

Different methods have been introduced for robust estimation of GARCH models. In this work, we compare our estimator to two recent ones that have been shown to outperform previous approaches. These are the recursive robust evaluation of parameters based on outlier criterion statistics of Charles and Darne (2005) and the robust GARCH model of Muler and Yohai (2008) based on a generalized class of M-estimators.

The literature usually distinguishes between two families of outliers: additive and innovative outliers. The former are characterized by single abnormal observations, whereas the latter have effects that propagate all along the process. Here, we consider additive outliers in the conditional volatility of the simple GARCH(1,1) model introduced by Bollerslev (1986). Note, however, that the proposed method can be applied to other GARCH models for which maximum likelihood estimation is possible.

The remainder of this paper is organized as follows. Section 2 recalls the definition of the GARCH model and its MLE. Further, the TLE is presented, followed by an introduction to the proposed GARCH WTLE. In section 3, we describe how the trimming parameter can be estimated. And section 4 presents the results of a Monte-Carlo simulation that compares the GARCH WTLE with recently introduced robust GARCH estimators. Conclusions and ideas for future work are offered in the last section.

2. WTL GARCH(p,q)

For a stationary time series $x_1, x_2, \ldots, x_t, \ldots, x_N$ with mean process $x_t = E(x_t|\Omega_{t-1}) + \varepsilon_t$ and innovation terms ε_t, the GARCH model introduced by

Bollerslev (1986) can be defined as

$$
\begin{aligned}
\varepsilon_t &= z_t \sigma_t \ , \\
z_t &\sim \mathcal{D}_\phi(0,1) \ , \\
&\text{and} \\
\sigma_t^2 &= \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \ . \qquad (1)
\end{aligned}
$$

Here, Ω_{t-1} is the information known at time $t-1$ where $t \in \mathbb{Z}$. $\mathcal{D}_\phi$ is the distribution of the innovations $\mathbf{z}$ with mean zero, variance one, and additional distributional parameters $\phi \in \mathbf{\Phi}^I \subset \mathbb{R}^I$, where $I \in \mathbb{N}$. For example, the additional distributional parameter of innovations distributed according to Student's t distribution would be the degree of freedom ν. $p \in \mathbb{N}$ and $q \in \mathbb{N}^*$ are the order of the GARCH and ARCH terms respectively. Sufficient conditions for the conditional variance σ_t to be possitive are $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\sum_i^p \alpha_i + \sum_i^q \beta_i < 1$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. When all $\beta_i = 0$, we have the ARCH model of Engle (1982).

Given the model in (1) and an observed univariate return series, the MLE can be used to fit the set of parameters $\theta = \{\omega, \alpha, \beta, \phi\} \in \mathbf{\Theta}^J \subset \mathbb{R}^J$, where $J = 1 + p + q + I$ and θ includes the parameters of the GARCH model and of the innovation distribution. The estimates of the MLE are defined by

$$\hat{\theta}_{MLE} := \arg \max_{\theta \in \mathbf{\Theta}^J} \mathcal{L}(\theta) \ , \qquad (2)$$

where the log-likelihood function is

$$\mathcal{L}(\theta) = \ln \prod_{t=1}^{N} \mathcal{D}(\varepsilon_t; \sigma_t, \theta) \ . \qquad (3)$$

In particular, expression (3) reduces to the so-called quasi-maximum likelihood estimator (QML) when the innovations are assumed to be normally distributed,

$$\mathcal{L}_{QML}(\theta) = -\frac{1}{2} \sum_{t=1}^{N} [\log(2\pi) + \ln(\sigma_t^2) + \frac{\varepsilon_t^2}{\sigma_t^2}]. \qquad (4)$$

To reduce the impact of outliers in MLE, Hadi and Luceño (1997) and Vandev and Neykov (1998) introduced the WTLE,

$$\hat{\theta}_{WTLE} := \arg\min_{\theta \in \mathbf{\Theta}^J} \frac{1}{k} \sum_{i=1}^{k} w_{v(i)} f(y_{v(i)}; \theta). \tag{5}$$

$f(y_{v(1)};\theta) \leqslant f(y_{v(2)};\theta) \leqslant \cdots \leqslant f(y_{v(N)};\theta)$ is ordered in ascending order for fixed parameters θ and for the permutation index $v(i)$ of $f(y_i;\theta) = -\ln \mathcal{D}(y_i;\theta)$ with probability density $\mathcal{D}$ and weights w_i. The key idea in (5) is to trim the $n-k$ points that are the most unlikely in the estimation of the likelihood function. The WTLE reduces to (i) the MLE when $k = N$, (ii) the TLE when $w_{v(i)} = 1$ for $i \in (1, \ldots, k)$ and $w_{v(i)} = 0$ otherwise, and (iii) the median likelihood estimator (Vandev and Neykov, 1993) when $w_{v(k)} = 1$ and $w_{v(i)} = 0$ for $i \neq k$. However, the implementation of the WTLE might not be feasible for large data sets due to its combinatorial nature. Let us denote by "k sub-sample" the sub-sample of likely values with index in a sub-set of $\{1, \ldots, N\}$ of length k. Equation (5) leads then to the problem of finding the k sub-sample that minimizes the estimator. To overcome this limitation, Neykov and Müller (2003) introduced the fast-TLE, which involves repeated iterations of a two-step procedure—a trial step followed by a refinement step. First, a k sub-sample is used to fit an initial estimate of the parameters. Then, these estimates are used to calculate the likelihood values of all points in the data set. Third, the order index of the $N-k$ points that are the most unlikely is used as a new trimming index. The process is repeated until the convergence criteria are reached. Neykov and Müller (2003) showed that the refinement step always yields estimates with an improved or equivalent estimator.

The WTLE can be defined for GARCH models by combining (1), (3), and (5). This gives the weighted trimmed log-likelihood function

$$\mathcal{L}_{WTLE}(\theta) = \frac{1}{k} \sum_{i=1}^{k} w_{v(i)} \ln \mathcal{D}(\tilde{\varepsilon}_{v(i)}; \tilde{\sigma}_{v(i)}, \theta) \tag{6}$$

where

$$\tilde{\sigma}_t^2 = \omega \ + \ \sum_{i=1}^{p} \alpha_i \tilde{\varepsilon}_{t-i}^2 \ + \ \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}^2 \ , \tag{7}$$

and

$$\tilde{\varepsilon}_t^2 = \begin{cases} \varepsilon_t^2 & \text{if } t \leq v(k) \\ E(\varepsilon_t^2 | \Omega_{t-1}) & \text{if } t > v(k) \ . \end{cases} \tag{8}$$

$E(\varepsilon_t|\Omega_{t-1})$ is the expected value of the innovations at time t given past information. Here, $\tilde{\varepsilon}_t$ at index $t > v(k)$ are replaced by their expected values in the conditional variance (7) to limit the impact of the outliers along the time-dependent variance. In practice, one can use $E(\varepsilon_t^2|\Omega_{t-1}) = \sigma_t^2 r_c(\varepsilon_t^2/\sigma_t^2)$ in (8) as defined by Muler and Yohai (2008),

$$r_c(u) = \begin{cases} u & \text{if } u \leq c \\ c & \text{if } u > c \ . \end{cases} \tag{9}$$

3. Estimation of Trimming Parameter

Here we describe how to automatically define the trimming parameter k in (6). First, we define the absolute deviation of the log-likely values (ADLLVs),

$$\begin{aligned} \mathcal{V}_t &= |\ln \mathcal{D}(\varepsilon_t; \sigma_t, \theta) - \ln \mathcal{D}(E(\varepsilon_t|\Omega_{t-1}); \sigma_t, \theta)| \\ &= |\ln \frac{\mathcal{D}(\varepsilon_t; \sigma_t, \theta)}{\mathcal{D}(E(\varepsilon_t|\Omega_{t-1}); \sigma_t, \theta)}|, \end{aligned} \tag{10}$$

where ε_t are the innovations; θ, the set of GARCH parameters; and Ω_{t-1}, the past information described in equations (1). In the particular case of normally distributed innovations, $\mathcal{V}_t$ reduces to the conditional coefficient of variation $\mathcal{V}_{QML,t} = \frac{x_t^2}{\sigma_t^2}$.

The trimming parameter k is determined using the following multi-step iterative procedure: We first compute $\mathcal{V}_t$ with the initial parameter values of GARCH models that can be set or estimated by a biased etimator. We then identify the index of the largest absolute deviation values that form a cluster of unlikely points (Figure 1). This new index is then used in the WTLE to obtain new estimates. These estimates are used to compute a new $\mathcal{V}_t$. We repeat the identification of unlikely events until the weighted trimmed log-likelihood function (6) reaches a maximum. Note that in practice, convergence is achieved in only a few steps.

4. Simulation Study

4.1. Software and Hardware Employed

All the models compared in this study were implemented in the R statistical programming language (R Development Core Team, 2010). The computation time reported is only an indication and may change with the platform. Nevertheless, note the purpose of this work was not to obtain the most efficient implementation. We used a 64-bit Darwin kernel running on an eight-core system with 16 GB of RAM. All routines were implemented in R, except the computation of the likelihood function, which was implemented in C (code available upon request).

4.2. Trimming Paramater k

To assess the efficiency of our approach to estimate the trimming parameter k, we performed a Monte-Carlo simulation of 1000 runs. For each run, we generated a GARCH(1,1) time series with parameters $\omega = 0.1$, $\alpha = 0.2$, and $\beta = 0.6$ as described in (1). Given the time series process x_t with conditional standard deviation σ_t, we defined a contaminated time series y_t. Outliers at time index i were defined as $y_i = d \cdot \sigma_i$, where d is the outlier scale and σ_i the conditional standard deviation of x_i at time i. For $i \neq j$, we have $y_j = x_j$. The length of the time series was 1500 and the levels of contamination were set to 1%, 5%, and 10% with outlier scales $d \in \{4, 6, 10\}$. The indices of the additive outliers were taken from a truncated Poisson distribution with truncation 10.

For large outlier scales, the identification of unlikely points converged to the correct index in most of the trials, whereas for smaller outliers, the approach

d	4											6		10
Offset	0	1	2	3	4	5	6	7	8	9	11	0	1	0
1%	573	274	94	34	12	7	4	1	.	1	.	993	7	1000
5%	617	248	76	33	15	5	2	2	1	1	.	991	9	1000
10%	676	211	71	20	12	6	2	1	.	.	1	997	3	1000

Table 1: Identification of unlikely points for 1000 simulation runs. d is the scale factor of the outliers. Positive offsets are counts of events that were superfluously identified as outliers. Negative offsets are counts of outliers that were not identified. Zero offset then means that all outliers and no other events have been identified as abnormal points. The length of the simulated GARCH(1,1) is 1000 with parameters $\omega = 0.1$, $\alpha = 0.2$, and $\beta = 0.6$

might consider superfluous points as unlikely (Table 1). However, the impact of considering few points as outliers in the GARCH WTLE has negligible impact on the estimates, as we will show in the results of the simulation study.

This method differs from the fast-TLE of Müller and Neykov (2003) because it affords an estimate of the trimming parameter k. However, the method does not guarantee the best trimming parameter. Indeed, the choice of the trimming parameter k is subject to the method used to identify the cluster of unlikely values. In practice, however, the approach yields promising results, as will be described in the next section.

4.3. Robust Estimation

We compared the GARCH WTLE to the quasi-maximum likelihood estimator (QMLE), the GARCH M-estimators (M1, M2) with their bounded versions (BM1, BM2) introduced by Muler and Yohai (2008), and the recursive robust GARCH estimator (REC) of Charles and Darne (2005). For the M1, M2, BM1, and BM2 estimators, we set the robust parameters to the values recommended by the authors. However, in the REC, we used stronger threshold statistics ($c = 4$) than recommended by Charles and Darne (2005). Indeed, we noticed that for large outliers, it is crucial to use a low threshold, for otherwise, the unfiltered outliers will lead to poor convergence rates for the optimization routines. The trimming parameter for our GARCH WTLE was automatically defined as described in section 3.

For a contaminated time series y_t of x_t, where $y_t = x_t$ for $t \neq i$ and with outliers $y_i = d \cdot \sigma_i$ at time index i and scale d, we compared the models on the basis of the mean square error (MSE) and relative mean square error (RMSE) of their estimates. The sample MSE of the contaminated series y_t can be defined as

$$\widehat{\text{MSE}}_y = \frac{1}{N} \sum_{i=1}^{N} (\hat{\theta}_i - \theta)^2 \ , \tag{11}$$

where $\hat{\theta}$ are the fitted parameters of the contaminated series y_t, and θ are the parameters used to simulate the series. N is the number of Monte-Carlo runs. We expressed the RMSE as the ratio of the MSE of the estimates of y_t with its counterpart for the uncontaminated series x_t,

$$\widehat{\text{RMSE}} = \frac{\widehat{\text{MSE}}_y(\theta_{WTLE})}{\widehat{\text{MSE}}_x(\theta_{MLE})} \ . \tag{12}$$

We considered 5000 runs of simulated GARCH(1,1) series of length 1500 as described in (1) with 1%, 5%, and 10% outliers. We used a range of outlier scales, $d \in \{3, 5, 10\}$, to study how the methods performs with large and small abnormal points. The outliers were taken from a truncated Poisson distribution with truncation 10. We also determined the percentage of optimization runs that converged to an optimal solution. We used two sets of parameters for the GARCH(1,1) model and explicitly set the starting values in the optimization routines different from the optimal values.

In the first study (Table 2), we used the parameters $\omega = 1$, $\alpha = 0.5$, and $\beta = 0.4$ to simulate the time series. Although these values are not typical estimates encountered with financial returns, i.e., a large β and small α with a persistence close to 1, we used them in order to compare our results with the estimates in (Muler and Yohai, 2008). In the second study (Table 3), we used more realistic coefficients: $\omega = 1$, $\alpha = 0.2$, and $\beta = 0.6$.

As a last comparison, we report the MSE in figure 2 for outlier scales ranging from 3 to 10. Again, we preformed 5000 simulation runs to estimate the MSE. The length of simulated series is 1500 and the GARCH parameters used are $\omega = 1$, $\alpha = 0.2$, and $\beta = 0.7$.

As noted in Charles and Darne (2005), the bounded M-estimators (BM1, BM2) have a smaller bias than the unbounded versions, but are subject to a lower rate of convergence. Moreover, M2 and BM2 have better estimates than the corresponding less-robust models, M1 and BM1 respectively. And although the estimates of the REC are close to those of the other estimators, the computation time was much larger than those of the others (Tables 2 and 3).

We note that throughout the simulation study, the choice of the robustness tuning parameters was crucial to obtaining reasonable estimates with the REC, M1, M2, BM1 and BM2. By contrast, our automatic method, introduced in the previous section, obviated any need for tedious parameter refinement. This feature can be particularly appealing to practitioners.

Overall, the WTLE yields estimates with the smallest MSE and RMSE and yet incurs only a small computational cost (Tables 2 and 3).

5. Conclusions and Future Work

We successfully applied the WTLE to GARCH modeling and showed through an extensive simulation study that it provides robust and reliable estimates with a small computation cost. Moreover, the proposed fully automatic method for selecting the trimming parameter obviates the tedious fine tuning process required by other models to obtain a "robust" parameter. We also note that only the simple GARCH(1,1) was used in this model. However, the WTLE can be used with any GARCH model for which there exists a likelihood estimator. We are currently exploring the applicability of the WTLE with multivariate GARCH models and have obtained promising preliminary results.

Acknowledgments

This work was made possible by financial support from Finance Online GmbH and from the Swiss Federal Institute of Technology in Zurich. We are grateful for the comments of the participants of the 4th Meielisalp Workshop and Summer School 2010 held in Meielisalp, Switzerland. We thank Stefano Iacus for a critical reading of the manuscript and for his helpful comments. Any remaining errors are ours alone.

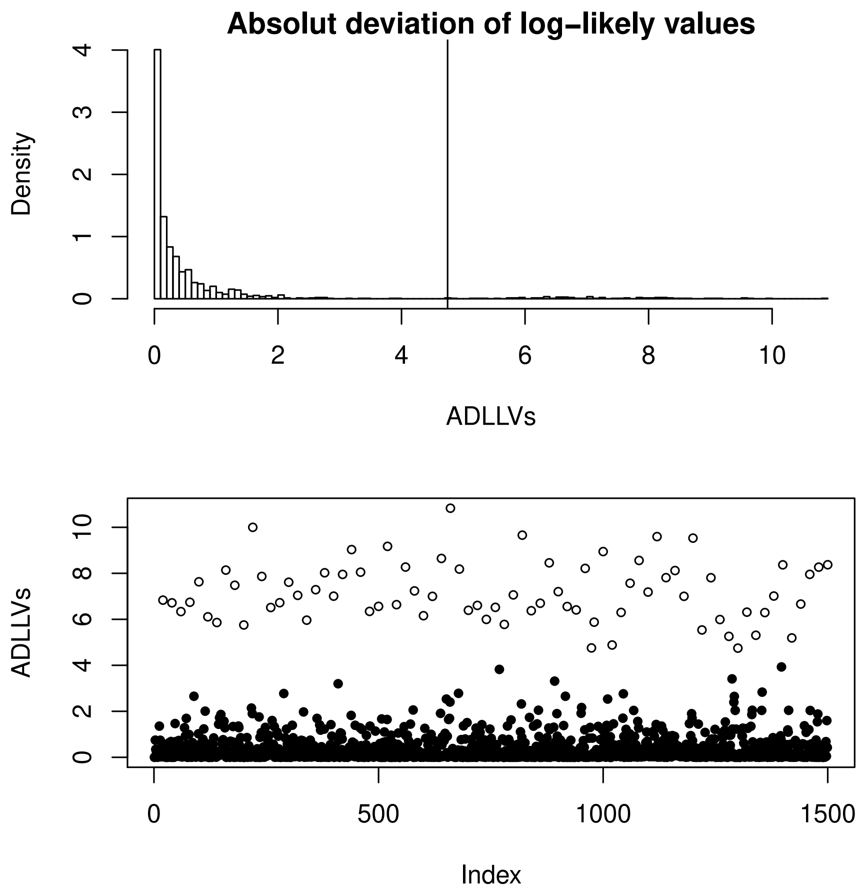

Figure 1: The upper figure is a histogram of the ADLLVs for a simulated GARCH(1,1) series of length 1500 with parameters $\omega = 0.1$, $\alpha = 0.2$, and $\beta = 0.6$ with 75 outliers of scale $d = 5$. The lower figure plots the ADLLVs, with the most unlikely values in empty circles.

1% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	0.27	0.01	0.00	100%	0.01	0.72	0.02	0.01	99%	0.01	9.22	0.09	0.03	51%	0.01
	RMSE	1.42	1.93	1.33			3.80	8.59	3.80			48.94	38.80	15.03		
REC	MSE	0.35	0.09	0.03	100%	2.73	0.29	0.08	0.02	100%	2.80	0.29	0.08	0.02	100%	2.82
	RMSE	1.88	37.13	14.10			1.55	35.26	13.18			1.57	35.78	13.66		
M1	MSE	0.35	0.01	0.01	98%	0.05	0.40	0.01	0.01	98%	0.05	1.09	0.01	0.03	97%	0.05
	RMSE	1.84	6.33	3.87			2.10	4.67	4.58			5.79	3.98	17.12		
BM1	MSE	0.28	0.01	0.01	90%	0.07	0.28	0.01	0.01	90%	0.07	0.27	0.01	0.01	90%	0.07
	RMSE	1.51	2.76	3.17			1.46	2.71	3.10			1.44	2.72	3.12		
M2	MSE	0.35	0.01	0.01	96%	0.06	0.52	0.01	0.01	97%	0.06	1.34	0.02	0.04	94%	0.06
	RMSE	1.86	6.52	5.61			2.77	6.03	7.81			7.09	6.78	24.37		
BM2	MSE	0.27	0.02	0.01	84%	0.08	0.26	0.02	0.01	83%	0.08	0.26	0.02	0.01	85%	0.08
	RMSE	1.41	6.98	3.37			1.37	6.92	3.28			1.39	7.00	3.31		
WTL	MSE	0.20	0.00	0.00	97%	0.36	0.18	0.00	0.00	98%	0.31	0.19	0.00	0.00	98%	0.31
	RMSE	1.05	1.75	1.55			0.98	1.01	1.06			1.01	1.04	1.05		

5% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	1.70	0.02	0.01	98%	0.01	23.61	0.09	0.05	63%	0.01	20.85	0.24	0.30	22%	0.03
	RMSE	8.92	6.55	4.88			123.77	38.19	29.66			109.28	98.76	166.60		
REC	MSE	0.98	0.12	0.04	99%	3.00	0.26	0.09	0.03	100%	3.68	0.30	0.09	0.03	99%	3.95
	RMSE	5.16	50.78	21.78			1.36	35.99	15.07			1.56	37.31	16.59		
M1	MSE	2.27	0.03	0.02	98%	0.06	1.85	0.04	0.05	98%	0.06	3.49	0.09	0.14	86%	0.07
	RMSE	11.91	13.50	9.59			9.68	17.32	27.42			18.31	39.12	81.54		
BM1	MSE	0.58	0.02	0.01	88%	0.08	0.28	0.02	0.01	89%	0.08	0.29	0.02	0.01	88%	0.08
	RMSE	3.06	7.61	4.78			1.49	6.86	4.37			1.51	6.66	4.37		
M2	MSE	0.93	0.04	0.02	98%	0.07	1.87	0.07	0.06	95%	0.08	3.48	0.09	0.13	75%	0.08
	RMSE	4.85	15.45	10.04			9.79	29.95	35.04			18.24	38.26	74.19		
BM2	MSE	0.45	0.01	0.01	81%	0.08	0.26	0.01	0.01	82%	0.08	0.26	0.01	0.01	83%	0.08
	RMSE	2.36	3.86	4.22			1.36	3.38	3.71			1.38	3.48	3.88		
WTL	MSE	0.51	0.01	0.01	98%	0.38	0.20	0.00	0.00	96%	0.31	0.20	0.00	0.00	96%	0.30
	RMSE	2.66	2.42	2.67			1.07	1.18	1.28			1.08	1.12	1.24		

10% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	4.94	0.02	0.02	96%	0.01	71.84	0.10	0.11	45%	0.02	24.62	0.23	0.30	38%	0.02
	RMSE	25.26	8.47	9.45			367.45	42.93	56.78			125.93	94.89	158.35		
REC	MSE	2.62	0.15	0.06	94%	3.04	0.19	0.08	0.03	100%	4.31	2.19	0.12	0.05	96%	4.52
	RMSE	13.39	62.34	32.22			0.99	34.23	16.75			11.19	48.16	27.59		
M1	MSE	6.71	0.07	0.03	97%	0.07	1.41	0.14	0.07	98%	0.07	3.70	0.12	0.14	84%	0.09
	RMSE	34.33	29.03	17.84			7.19	55.25	37.57			18.93	50.07	74.53		
BM1	MSE	1.58	0.03	0.01	86%	0.08	0.25	0.03	0.01	86%	0.08	0.26	0.03	0.01	86%	0.08
	RMSE	8.08	14.03	6.73			1.27	11.32	4.92			1.33	10.86	4.71		
M2	MSE	2.38	0.08	0.03	98%	0.07	0.99	0.18	0.07	97%	0.09	3.51	0.11	0.14	70%	0.09
	RMSE	12.19	33.12	15.11			5.04	72.23	39.34			17.96	46.08	72.99		
BM2	MSE	1.35	0.01	0.01	81%	0.08	0.24	0.01	0.01	81%	0.08	0.26	0.01	0.01	81%	0.08
	RMSE	6.92	5.13	5.98			1.25	2.87	3.95			1.30	2.84	4.14		
WTL	MSE	0.86	0.01	0.01	100%	0.37	0.25	0.00	0.00	99%	0.30	0.23	0.00	0.00	99%	0.30
	RMSE	4.41	2.75	3.13			1.28	1.36	1.36			1.15	1.20	1.35		

Table 2: Mean square error and relative mean square error for the simple GARCH(1,1) with 1%, 5% and 10% outliers, $y_i = d{\cdot}\sigma_i$, with scale $d = \{3, 5, 10\}$ and parameters $\omega = 1$, $\alpha = 0.5$, and $\beta = 0.4$. The length of simulated series is 1500 and the number of Monte-Carlo replications is 5000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

1% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	0.16	0.00	0.01	100%	0.02	1.49	0.01	0.04	99%	0.01	14.95	0.13	0.13	37%	0.01
	RMSE	2.19	1.27	1.40			20.11	5.14	7.47			201.79	117.43	27.31		
REC	MSE	0.23	0.01	0.02	99%	1.69	0.24	0.01	0.02	98%	1.63	0.24	0.01	0.02	98%	1.66
	RMSE	3.03	10.35	4.74			3.19	10.13	4.92			3.22	10.54	5.02		
M1	MSE	0.17	0.00	0.01	96%	0.06	0.54	0.00	0.03	98%	0.06	3.21	0.01	0.16	80%	0.07
	RMSE	2.28	1.94	2.13			7.25	3.49	5.65			43.38	8.14	33.20		
BM1	MSE	0.24	0.00	0.02	92%	0.08	0.24	0.00	0.02	92%	0.08	0.24	0.00	0.02	92%	0.08
	RMSE	3.20	2.73	3.39			3.22	2.72	3.38			3.28	2.72	3.44		
M2	MSE	0.32	0.00	0.02	93%	0.07	0.81	0.01	0.05	95%	0.08	3.02	0.01	0.17	76%	0.08
	RMSE	4.33	3.15	4.20			10.95	5.37	9.33			40.80	11.93	34.11		
BM2	MSE	0.23	0.01	0.02	88%	0.08	0.23	0.01	0.02	89%	0.08	0.24	0.01	0.02	87%	0.08
	RMSE	3.17	6.86	3.66			3.17	6.88	3.68			3.22	6.93	3.71		
WTL	MSE	0.07	0.00	0.01	97%	0.38	0.08	0.00	0.01	98%	0.33	0.08	0.00	0.01	98%	0.33
	RMSE	1.00	1.15	1.21			1.03	1.02	1.04			1.04	1.01	1.03		

5% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	3.79	0.00	0.04	99%	0.01	43.56	0.02	0.17	12%	0.02	135.92	0.04	0.09	0%	0.03
	RMSE	51.92	3.07	8.88			595.98	19.21	35.04			1859.42	33.35	18.22		
REC	MSE	0.23	0.01	0.03	98%	3.07	0.30	0.02	0.03	97%	2.93	0.30	0.01	0.03	96%	3.04
	RMSE	3.19	13.28	5.53			4.17	14.17	6.70			4.13	13.19	6.80		
M1	MSE	0.62	0.01	0.02	93%	0.07	0.83	0.03	0.05	92%	0.08	2.74	0.04	0.08	55%	0.09
	RMSE	8.50	9.90	4.94			11.40	25.73	10.76			37.52	33.19	15.93		
BM1	MSE	0.25	0.00	0.02	91%	0.09	0.25	0.00	0.02	91%	0.08	0.29	0.00	0.02	91%	0.08
	RMSE	3.47	3.71	3.82			3.47	3.65	3.77			3.99	3.20	3.86		
M2	MSE	0.46	0.01	0.03	94%	0.08	1.02	0.03	0.06	83%	0.10	3.66	0.04	0.12	48%	0.11
	RMSE	6.34	10.71	6.29			13.90	25.55	11.80			50.03	31.89	24.97		
BM2	MSE	0.24	0.01	0.02	88%	0.09	0.24	0.01	0.02	87%	0.08	0.28	0.01	0.02	88%	0.08
	RMSE	3.26	4.04	3.68			3.29	4.09	3.70			3.86	4.22	4.04		
WTL	MSE	0.08	0.00	0.01	97%	0.40	0.08	0.00	0.01	96%	0.32	0.08	0.00	0.01	96%	0.32
	RMSE	1.12	1.35	1.37			1.08	1.13	1.11			1.07	1.10	1.09		

10% with		$d=3$					$d=5$					$d=10$				
		ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]	ω	α	β	conv.	time [s]
QML	MSE	13.36	0.01	0.10	90%	0.02	28.69	0.03	0.13	2%	0.03	4.22	0.04	0.12	0%	0.03
	RMSE	170.38	6.97	20.77			365.73	27.73	25.02			53.83	32.06	24.25		
REC	MSE	0.44	0.02	0.03	37%	3.88	0.40	0.02	0.05	92%	3.79	0.38	0.02	0.04	93%	3.92
	RMSE	5.56	14.71	6.52			5.07	17.31	9.04			4.86	15.18	8.68		
M1	MSE	0.85	0.02	0.03	65%	0.07	0.70	0.03	0.05	90%	0.08	3.48	0.04	0.10	51%	0.09
	RMSE	10.82	13.33	6.09			8.91	28.03	10.53			44.38	32.43	20.50		
BM1	MSE	0.27	0.01	0.02	92%	0.08	0.28	0.01	0.02	92%	0.08	0.32	0.00	0.02	91%	0.08
	RMSE	3.46	4.70	3.89			3.52	4.55	3.82			4.08	3.83	3.87		
M2	MSE	0.48	0.02	0.04	94%	0.09	0.90	0.03	0.06	80%	0.10	4.17	0.04	0.12	49%	0.11
	RMSE	6.18	14.88	6.88			11.50	27.51	11.49			53.12	32.96	24.58		
BM2	MSE	0.25	0.00	0.02	87%	0.08	0.26	0.00	0.02	88%	0.08	0.29	0.00	0.02	87%	0.08
	RMSE	3.21	3.31	3.62			3.27	3.31	3.69			3.73	3.46	4.07		
WTL	MSE	0.09	0.00	0.01	100%	0.39	0.09	0.00	0.01	99%	0.32	0.09	0.00	0.01	99%	0.31
	RMSE	1.18	1.47	1.47			1.15	1.18	1.19			1.15	1.17	1.19		

Table 3: Mean square error and relative mean square error for the simple GARCH(1,1) with 1%, 5% and 10% outliers, $y_i = d{\cdot}\sigma_i$, with scale $d = \{3, 5, 10\}$ and parameters $\omega = 1$, $\alpha = 0.2$, and $\beta = 0.6$. The length of simulated series is 1500 and the number of Monte-Carlo replications is 5000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

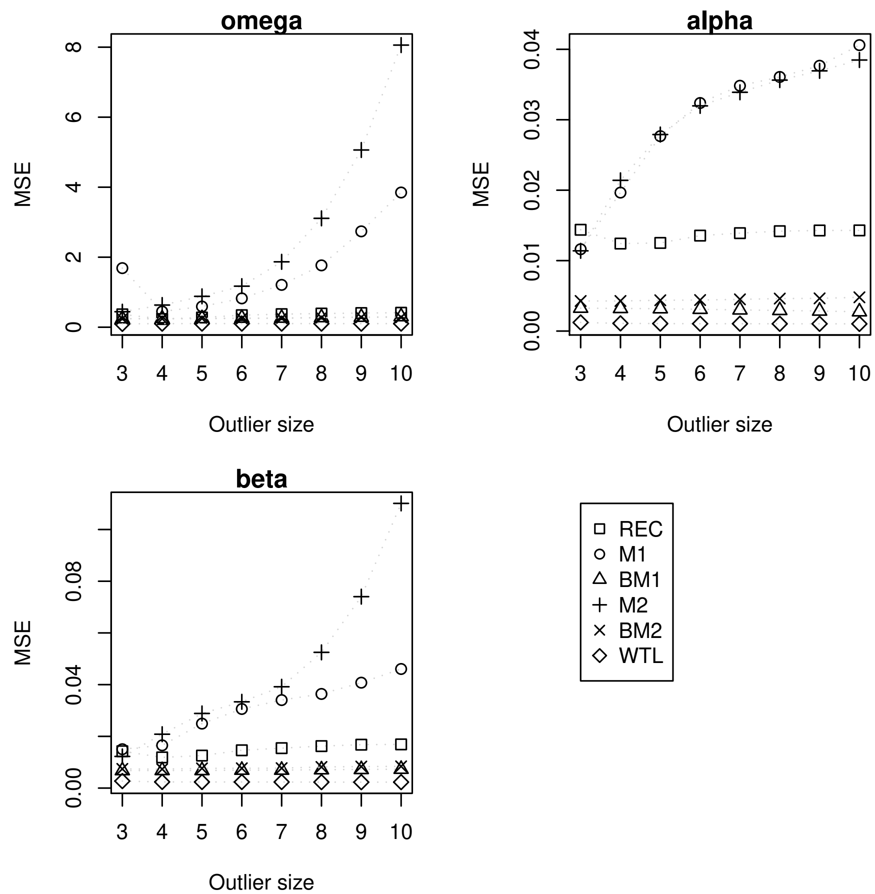

Figure 2: MSE for the simple GARCH(1,1) with 5% outliers, $y_i = d{\cdot}\sigma_i$, and with scale d ranging from 3 to 10. GARCH parameters used are $\omega = 1$, $\alpha = 0.2$, and $\beta = 0.7$. The length of simulated series is 1500 and the number of Monte-Carlo replications is 5000.

References

Bednarski, T., Clarke, B., 1993. Trimmed likelihood estimation of location and scale of the normal distribution. Australian Journal of Statistics 35, 141–153.

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

Bollerslev, T., 2008. Glossary to arch (garch) .

Brooks, C., Burke, S., Persand, G., 2001. Benchmarks and the accuracy of garch model estimation. International Journal of Forecasting 17, 45 – 56.

Charles, A., Darne, O., 2005. Outliers and garch models in financial data. Economics Letters 86, 347–352.

Cizek, P., 2008. Greneral trimmed estimation: robust approach to nonlinear and limited dependent variable models. Econometric Theory 24, 1500–1529.

Dimova, R., Neykov, N., 2004. Generalized d-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications. Theory and applications of recent robust methods , 83–91.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50, 987–1007.

Hadi, A., Luceño, A., 1997. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. Computational Statistics & Data Analysis 25, 251 – 272.

Hotta, L., Tsay, R., 1998. Outliers in garch processes.

Markatou, M., 2000. Mixture models, robustness, and the weighted likelihood methodology. Biometrics 56, 483–486.

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

Muler, N., Yohai, V., 2008. Robust estimates for garch models. Journal of Statistical Planning and Inference 138, 2918 – 2940.

Müller, C., Neykov, N., 2003. Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. Journal of Statistical Planning and Inference 116, 503 – 519.

Neykov, N., Filzmoser, P., Dimova, R., Neytchev, P., 2007. Robust fitting of mixtures using the trimmed likelihood estimator. Computational Statistics & Data Analysis 52, 299 – 308.

Neykov, N., Müller, C., 2003. Breakdown point and computation of trimmed likelihood estimators in generalized linear models, in: Developments in robust statistics: International Conference on Robust Statistics 2001, Physica Verlag. p. 277.

R Development Core Team, 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Vandev, D., Neykov, N., 1993. New directions in statistical data analysis and robustness. Birkhauser Verlag. chapter Robust maximum likelihood in the gaussian case. pp. 257 – 264.

Vandev, D., Neykov, N., 1998. About regression estimators with high breakdown point. Statistics 32, 111–129.