ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Robust Estimation with the Weighted Trimmed Likelihood Estimator

Yohan Chalabi and Diethelm Würtz

No. 2010-01

ETH Econohysics Working and White Papers Series
Online at https://www.rmetrics.org/WhitePapers

Robust Estimation with the Weighted Trimmed Likelihood Estimator

Yohan Chalabi* Diethelm Wuertz
Institute for Theoretical Physics, ETH Zurich, Switzerland
Computational Science and Engineering, ETH Zurich, Switzerland

November 2012

We consider the problem related to the estimation of parametric models in the presence of outliers. The maximum likelihood estimator is often used to find parameter values. However, it is highly sensitive to abnormal points. In this regard, the weighted trimmed likelihood estimator (WTLE) has been introduced as a robust alternative. We present a new scheme for automatically computing the trimming parameter and weights of the WTLE. The method is illustrated by applying it to the standard GARCH model. We compare the approach with other recently introduced robust GARCH estimators through an extensive simulation study.

Keywords GARCH models · Robust estimators · outliers · Weighted trimmed likelihood estimator (WTLE) · Quasi maximum Likelihood estimator (QMLE)

JEL Classification C13 · C22

1 Introduction

We consider the problem of estimating the parameters of the stochastic models that are widely used in econometrics. One of the most important models is the generalized autoregressive conditional heteroskedasticity model (GARCH), which is used to model the volatility clustering of financial returns. GARCH originates in the ARCH model of Engle (1982), who won the

*Corresponding author. Email address: chalabi@phys.ethz.ch. Postal address: Institut für Theoretische Physik, HIT G 31.5, Wolfgang-Pauli-Str. 27, 8093 Zürich, Switzerland.

Nobel-memorial prize "for methods of analyzing economic time series with time-varying volatility (ARCH)". The family of GARCH-type models is often used as the foundation of risk measures.

The traditional approach for estimating parameter values of a model such that it is well fitted to a given data set, is to apply the maximum likelihood estimator (MLE). Unfortunately, the MLE can be highly sensitive to any outliers that might be present in the data. This is especially problematic when constructing a risk measure that is reliant upon a distribution model, since the estimated parameters may acquire a bias from the estimator.

The estimation of distributional model parameters in the presence of abnormal points is an active field of research known as robust statistics. The aim of this chapter is to improve the recently introduced weighted trimmed likelihood estimator (WTLE) in order to provide a viable alternative to the maximum likelihood estimator (MLE). To this end, a scheme for automatically computing the parameters of the WTLE estimation is introduced. This auto-WTLE is used to obtain robust estimates of the GARCH parameters. The performance of this new approach is compared to that of other robust GARCH models, in an extensive Monte-Carlo simulation.

The remainder of this paper is organized as follows. Section 2 recalls the maximum likelihood estimator and its sensitivity to outliers, and then presents the WTLE. Section 3 introduces the automatic approach to calculating the weights and the trimming parameter that are required for the WTLE. The Auto-WTLE is applied to the GARCH model in Section 4, and its performance is tested against that of the recently introduced robust GARCH models. Concluding remarks, speculation and ideas are presented at the end.

2 The weighted trimmed likelihood estimator

The maximum likelihood estimator (MLE), introduced by Fisher (1922), is used to relate parameters of the probability distributions in a statistical model of some data set, to the likelihood of observing the outcomes. Suppose a random sample of n *iid* observations, $\{x_1, x_2, \ldots, x_n\}$, is drawn from an unknown continuous probability density distribution, f_θ, with parameters θ. The estimated parameters, $\hat{\theta}$, are the set of parameters that are most likely, given some assumed model,

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \boldsymbol{\Theta}} \prod_{i=1}^{n} f_\theta(x_i),$$

where $\boldsymbol{\Theta}$ is the set of feasible parameters for f.

The above maximization problem is usually transformed to the equivalent log-likelihood, problem which can be expressed in terms of a sum rather than a product. The parameters $\hat{\theta}$ can then be estimated by the maximum log-likelihood function,

$$\mathcal{L}_{\text{MLE}}(\theta) = \sum_{i=1}^{n} \ln f_\theta(x_i). \tag{1}$$

It is well known that the MLE is highly sensitive to outliers. This can be understood through a geometrical interpretation. The MLE in Eq. (1) is equivalent to

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta \in \mathbf{\Theta}} \ \mathcal{L}_{\text{MLE}}(\hat{\theta}) \\ &= \arg\max_{\theta \in \mathbf{\Theta}} \ \frac{1}{n} \sum_{i=1}^{n} \ln f_\theta(x_i) \\ &= \arg\max_{\theta \in \mathbf{\Theta}} \ \ln \sqrt[n]{f_\theta(x_1)\, f_\theta(x_2)\, \ldots\, f_\theta(x_n)} \\ &= \arg\max_{\theta \in \mathbf{\Theta}} \ \sqrt[n]{f_\theta(x_1)\, f_\theta(x_2)\, \ldots\, f_\theta(x_n)}.\end{aligned}$$

The MLE is therefore equivalent to maximizing the geometric mean of the likely values. This maximum is attained when each of the geometric mean's arguments are of equal size. In other words, the MLE yields estimates for which the likely values are most equal. In the presence of outliers—i.e., in the presence of values for which the likely values in the assumed distributional model are very small—the MLE would yield estimates that reduce the likeliness of the good data points in favor of obtaining more equally likely values across all events. A geometrical interpretation of the objective function can be made with other types of optimization; say, in the maximum product of spacing estimator introduced by Cheng and Amin (1983), and Ranneby (1984).

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

$$\hat{\theta}_{\text{WTLE}} = \arg\min_{\theta \in \mathbf{\Theta}} \frac{1}{k} \sum_{i=1}^{k} w_{v(i)}\, g_\theta(x_{v(i)}), \tag{2}$$

where $g_\theta(x_{v(1)}) \leqslant g_\theta(x_{v(2)}) \leqslant \cdots \leqslant g_\theta(x_{v(N)})$ are indexed in ascending order for fixed parameters θ and with permutation index $v(i)$ of $g_\theta(x_i) = -\ln f_\theta(x_i)$, f_θ is the probability density, and the w_i are weights. The key idea in Eq. (2) is to trim the $n-k$ points that are the most unlikely from the estimation of the likelihood function. The WTLE reduces to: (i) the MLE when $k = N$, (ii) the trimmed likelihood estimator when $w_{v(i)} = 1$ for $i \in (1, \ldots, k)$ and $w_{v(i)} = 0$ otherwise, and (iii) the median likelihood estimator, as reported by Vandev and Neykov (1993), when $w_{v(k)} = 1$ and $w_{v(i)} = 0$ for $i \neq k$.

The WTLE is a generalization of the trimmed likelihood estimator (TLE) of Neykov and Neytchev (1990), see also the work of Bednarski and Clarke (1993), and Vandev and Neykov (1993). The WTLE has been applied to many different fields: Markatou (2000) used the weighted likelihood estimating equations 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. Vandev and Neykov (1998); Müller and Neykov (2003); Dimova and Neykov (2004) give a derivation of the breakdown point of the WTLE for various models.

The WTLE might can become unfeasible for large data sets, due to its combinatorial nature. Denote by "k sub-sample" the sub-sample of likely values with index i in a sub-set of length k among the full index set $\{1, \dots, N\}$.

Equation (2) then leads to the problem of finding the k sub-sample that minimizes the estimator. To avoid the optimization of this combinatorial problem, 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 make an initial estimate of the parameters. These estimates are then used to calculate the likelihood values of all points in the data set. Third, the order index of the least likely $N - k$ points is used as a new trimming index. This process is repeated until the convergence criteria are satisfied. Neykov and Müller (2003) showed that the refinement step always yields estimates with an improved or equivalent estimator value.

3 The auto-WTLE algorithm

In practice, a fixed value must be chosen for the trimming parameter in Eq. (2). If the trimming parameter, k, is too small, then the estimator might yield bad estimates due to the sensitivity of the MLE to any outliers that may be present in the data set. If, instead, the chosen trimming parameter is too large, this might result in biased estimates.

This section presents a new method for automatically selecting the trimming parameter, k, and the weights, ω_i, of the WTLE (Eq. 2). This method is a multi-step iterative procedure. The advantage of the new approach is that it obviates the fine-tuning process necessary to obtain a "robust" parameter in other models.

To start with, an initial set of parameters, θ, is chosen for the distributional model. These initial values could be chosen to be typical values for the given problem, or else could be obtained from another estimator. The probabilities, F, of the data points are then calculated in the assumed distributional model with parameters θ. As a consequence of the probability integral transform, if X are random variates with cumulative distribution function F, the probabilities $U = F(X)$, are uniformly distributed on $(0, 1)$. The spacings of the probabilities U then give an indication as to which points should be trimmed from, or down-weighted in, the WTLE.

Order statistics and order spacings are the foundations of non-parametric estimation. Goodness-of-fit tests are a good example of such non-parametric estimators. Now, define the spacings, D_t, as the differences between the consecutive ordered statistics $U_{(1)}$, $U_{(2)}$, $\dots$, $U_{(n)}$, where $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ with $U_0 = 0$ and $U_{n+1} = 1$. As noted by Pyke (1965), uniform spacings are interchangeable random variates. The distribution of D_i for any i matches the distribution of the first spacing D_1. Moreover, the distribution of the first spacing is, by definition, the same as F_{U_1}; the distribution of the first order statistics of the uniform distribution over the interval $(0, 1)$. As described by David and Nagaraja (2003), the cumulative distribution function of the order statistics, $X_{(1)}$, is given by

$$\begin{aligned} F_{X_{(1)}}(x) &= Pr\{X_{(1)} \leq x\} \\ &= 1 - Pr\{X_{(1)} > x\} \\ &= 1 - Pr\{\text{all } X_t > x\} \\ &= 1 - [1 - F_X(x)]^n . \end{aligned}$$

where n is the sample size. The distribution of uniform spacings becomes

$$F_{D_t}(x) = F_{D_1}(x) = F_{U_1}(x) = 1 - (1 - x)^n.$$

Hence, the probability distribution function of the uniform spacings is

$$f_{D_t}(x) = n(1 - x)^{n-1}.$$

Given that their theoretical distribution is known, the spacings, D_i, can now be filtered by a Monte-Carlo regime-switching model to determine the probabilities of the spacings, D_t, to be incorporated within the regime characterized by the density probability $f_{D_t} = n(1 - x)^{n-1}$. In addition to the theoretical regime of the spacings, two alternative regimes are considered. One of these has a density probability that corresponds to a larger sample size, $m_1 = a\,n$ with $a > 1$, while the other corresponds to a smaller sample size, $m_2 = b\,n$ with $0 < b < 1$. This results in a mixture model with three components:

$$f_{D_t}(x) \begin{cases} n(1 - x)^{n-1} & \text{if } s_t = 0, \\ m_1(1 - x)^{m_1-1} & \text{if } s_t = 1, \\ m_2(1 - x)^{m_2-1} & \text{if } s_t = 2, \end{cases}$$

where s_t is the state regime variable of the Markov chain with transition matrix,

$$P = \begin{bmatrix} p_{ij} \end{bmatrix},$$

where $p_{ij} = \mathbb{P}(s_t = i \mid s_{t-1} = j)$ is the transition probability from state regime j to state regime i. Here, a simplified transition matrix

$$\begin{bmatrix} p_{00} & \frac{1}{2}(1 - p_{00}) & \frac{1}{2}(1 - p_{00}) \\ \frac{1}{2}(1 - p_{00}) & p_{00} & \frac{1}{2}(1 - p_{00}) \\ \frac{1}{2}(1 - p_{00}) & \frac{1}{2}(1 - p_{00}) & p_{00} \end{bmatrix}$$

is used, which gives good results in practice as will be seen in the empirical study in Section 4.

The probabilities of being in the state $s_t = 0$ at time t are calculated by the scheme presented in (Kuan, 2002). Let $\mathcal{D}^t = \{d_1, d_2, \dots, d_t\}$ denote the collection of sample probability spacings up to index t. The conditional density of d_t, given information at $t-1$, $\mathcal{D}^{t-1}$, for the simplified model is

$$\begin{aligned} f(d_t \,|\, \mathcal{D}^{t-1}) = n(1-d_t)^{n-1}&\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1}) \\ &+ \frac{1}{2} m_1 (1-d_t)^{m_1-1}[1 - \mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})] \\ &+ \frac{1}{2} m_2 (1-d_t)^{m_2-1}[1 - \mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})]. \end{aligned}$$

By the Bayes rule, the posterior probability of s_t being in state 0 is given by

$$\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t) = \frac{n(1-d_t)^{n-1}\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^{t-1})}{f(d_t \,|\, \mathcal{D}^{t-1})},$$

with the prior probability of s at time $t+1$ given information at time t, being,

$$\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t) = p_{00}\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t) + \frac{1}{2}(1-p_{00})[1 - \mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t)].$$

The probabilities of being in state $s_t = 0$ can therefore be obtained by solving the recursive system formed by the previous three equations.

Further, the probabilities $\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)$ can be smoothed to reduce the magnitude of abrupt regime switches. As recommended by Kuan (2002), the method of Kim (1994), which gives the smoothed probabilities, was used. Here, this results in

$$\mathbb{P}(s_t = 0 \,|\, \mathcal{D}^T) = \mathbb{P}(s_t = 0 \,|\, \mathcal{D}^t) \left[\frac{p_{00}\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^T)}{\mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)} + \frac{(1-p_{00})\left[1 - \mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^T)\right]}{1 - \mathbb{P}(s_{t+1} = 0 \,|\, \mathcal{D}^t)} \right].$$

The advantage of this simplified Markov chain model, is that there is no need to estimate its parameters by numerical optimization. Instead, only typical values must be provided for the alternative regimes, m_1 and m_2. The values $m_1 = 10n$ and $m_2 = m/10$, lead to good results in practice. However, the probability of being in state $s_k = 0$ at the starting index k, also needs to be specified. In this regards, the spacing at index $\lceil i/n \rceil$ is taken as the starting position because the data point that corresponds to the median of the data set can be expected to be described by the stochastic model under consideration.

The smoothed probabilities of the spacings, D_t, in the state assumed by their theoretical distribution are then used as the weights in the WTLE (Eq. 2). From this, new set of parameters, θ^+, are obtained. The procedure is then repeated until the optimized weighted trimmed log-likelihood function reaches a maximum. In practice, it is sufficient to stop the procedure when the objective function has not been improved by more than a factor of 1% from the previous procedure step. This convergence is usually achieved within a few steps.

4 Robust GARCH modeling

Generalized autoregressive conditional heteroskedasticity (GARCH) models are widely used to reproduce the stylized facts of financial time series. Today, they play an essential role in risk management and volatility forecasting. It is therefore crucial to develop robust estimators for the GARCH. This section shows how to overcome this limitation by applying the robust weighted trimmed likelihood estimator (WTLE) to the standard GARCH model. The approach is compared with other recently introduced robust GARCH estimators. The results of an extensive simulation study subsequently show that the proposed estimator provides robust and reliable estimates with a small computation cost.

4.1 Introduction

Because time-variation of the volatility is a characteristic feature of financial time series, accurate modeling of this variation is critical in many financial applications. It is especially important in risk management. Since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model by Engle (1982), and of its generalization, the GARCH model, by Bollerslev (1986), copious theoretical and applied research work has been performed concerning these models. The success of the GARCH model and its derivations stem mainly from their ability to reproduce the typical properties exhibited by financial time series, particularly, volatility clustering, the fat-tailed return distributions, 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 (2009) compiled a glossary of more than 150 GARCH models. GARCH modeling is now a common practice, despite the fact that estimation of its parameters involves solving a rather difficult constrained nonlinear optimization problem. Moreover, it is common for different software implementations to produce conflicting estimates (Brooks et al., 2001). Besides the difficulty in parameter estimation, GARCH models remain, as do any other models, 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. These can result in abnormal points, or outliers, that are beyond the scope of the model. The maximum likelihood estimator (MLE) for GARCH models is very sensitive to these outliers, as was shown by Mendes (2000); Hotta and Tsay (1998).

Few different methods have been introduced for the robust estimation of GARCH model parameters. Two recent estimators that have been shown to outperform earlier approaches are the recursive robust evaluation of parameters based on outlier criterion statistics (Charles and Darne, 2005), and the robust GARCH model based on a generalized class of M-estimators (Muler and Yohai, 2008). These methods will be compared with a new estimator introduced here.

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

The remainder of this chapter is organized as follows. Section 4.2 first recalls the definition of the GARCH model and its MLE, and then presents the proposed GARCH WTLE. Then, in Section 4.3, the auto-WTLE algorithm presented in Section 3 is compared with the recently introduced robust GARCH estimators in an extensive Monte-Carlo simulation.

4.2 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) model the innovations as

$$\varepsilon_t = z_t \sigma_t, \tag{3a}$$

$$z_t \sim \mathcal{D}_\phi(0, 1), \tag{3b}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{3c}$$

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. The order of the ARCH and GARCH terms are $q \in \mathbb{N}^*$ and $p \in \mathbb{N}^*$, respectively. Sufficient conditions for the GARCH model to be stationary are $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, \ldots, q$, $j = 1, \ldots, p$, and $\sum_i^q \alpha_i + \sum_j^p \beta_j < 1$. When all $\beta_j = 0$, the GARCH model reduces to the ARCH model of Engle (1982).

Assuming the model in Eq. (3), and given an observed univariate financial return series, the MLE can be used to fit the set of parameters $\theta = \{\alpha, \beta, \phi\} \in \mathbf{\Theta}^J \subset \mathbb{R}^J$, where $J = 1 + p + q + I$ and θ includes the parameters of both the GARCH model and innovation distribution. The estimates of the MLE are defined by

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

where the log-likelihood function is

$$\mathcal{L}_{\text{MLE}}(\theta) = \ln \prod_{t=1}^{N} \mathcal{D}_\phi(\varepsilon_t, \sigma_t). \tag{4}$$

Equation (4) reduces to the so-called quasi-maximum likelihood estimator (QML) when the innovations are assumed to be normally distributed;

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

The WTLE can be defined for GARCH models by combining Eqs. (2) to (4). The estimates of the WTLE becomes

$$\hat{\theta}_{\text{WTLE}} = \arg \max_{\theta \in \mathbf{\Theta}^J} \frac{1}{k} \sum_{i=1}^{k} w_{v(i)} \ln \mathcal{D}_\phi(\tilde{\varepsilon}_{v(i)}, \tilde{\sigma}_{v(i)}),$$

where

$$\tilde{\sigma}_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \tilde{\varepsilon}_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2,$$

and $\mathcal{D}_\phi(\varepsilon_{v(1)}) \geq \mathcal{D}_\phi(\varepsilon_{v(2)}) \geq \cdots \geq \mathcal{D}_\phi(\varepsilon_{v(N)})$ is in descending order with permutation index $v(i)$. However, due to the recursive nature of the GARCH model, care must be taken to ensure that the unlikely innovations are not propagated through the conditional variance. The innovations, ε, are therefore reformulated in terms of their expected values when they are considered unlikely:

$$\tilde{\varepsilon}_t^2 = \begin{cases} \omega_t \varepsilon_t^2 + (1 - \omega_t)\mathbb{E}[\varepsilon_t^2 | \Omega_{t-1}] & \text{if } t \leq v(k), \\ \mathbb{E}[\varepsilon_t^2 | \Omega_{t-1}] & \text{if } t > v(k). \end{cases}$$

Note that the expected value of the squared innovations at time t given past information Ω_{t-1} corresponds to the conditional variance at time t, $\mathbb{E}[\varepsilon_t^2 | \Omega_{t-1}] = \sigma_t^2$, due to the definition of the distribution of the innovations, ε_t, in Eq. (3). This gives

$$\tilde{\varepsilon}_t^2 = \begin{cases} \omega_t \varepsilon_t^2 + (1 - \omega_t)\tilde{\sigma}_t^2 & \text{if } t \leq v(k), \\ \tilde{\sigma}_t^2 & \text{if } t > v(k). \end{cases}$$

Figures 1 and 2 illustrates the ability of the auto-WTLE to identify unlikely events for GARCH models. For large outlier scales, the identifier of unlikely points converged to the correct index (Fig. 1), whereas for smaller outliers, the approach might consider superfluous points to be unlikely (Fig. 2). However, as will be seen in Section 4.3, the impact of mistakenly considering few normal points to be outliers in the GARCH WTLE has negligible impact on the final parameter-value estimates. Figure 2 nicely illustrates one of the key advantages of the Auto-WTLE; this approach can identify values that appear too frequently than they ought to for the sample size. Such values are called inner outliers.

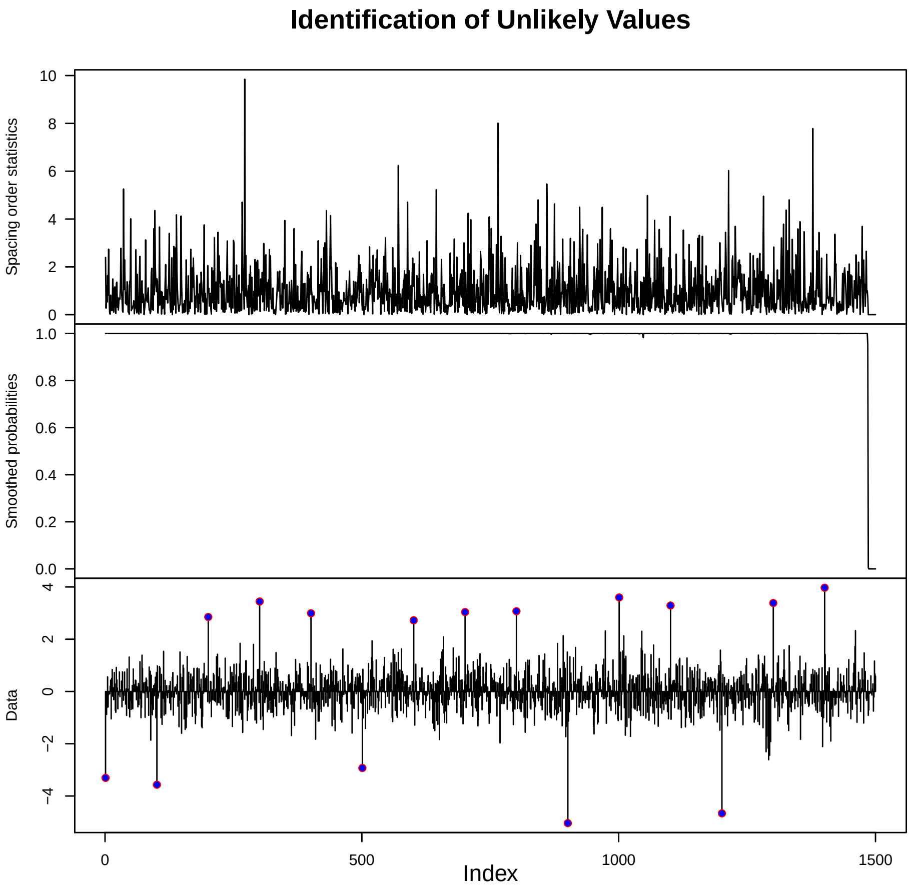

Figure 1: Estimation of a contaminated GARCH time series with the auto-WTLE estimator. The GARCH(1,1) series is of length `1500` with parameters $\omega = \texttt{0.1}$, $\alpha = \texttt{0.2}$, and $\beta = \texttt{0.6}$ with `15` equidistant outliers of scale d = `5`. The upper figure plots is the spacing order statistics. The middle figure displays the smoothed probabilites after applying the Mone-Carlo switching model filter on the spacings. The lower figure shows the time series. The empty circles are the points for which their likely values have been trimmed in the WTLE and the full cirlces are the exact outliers that were added to the time series.

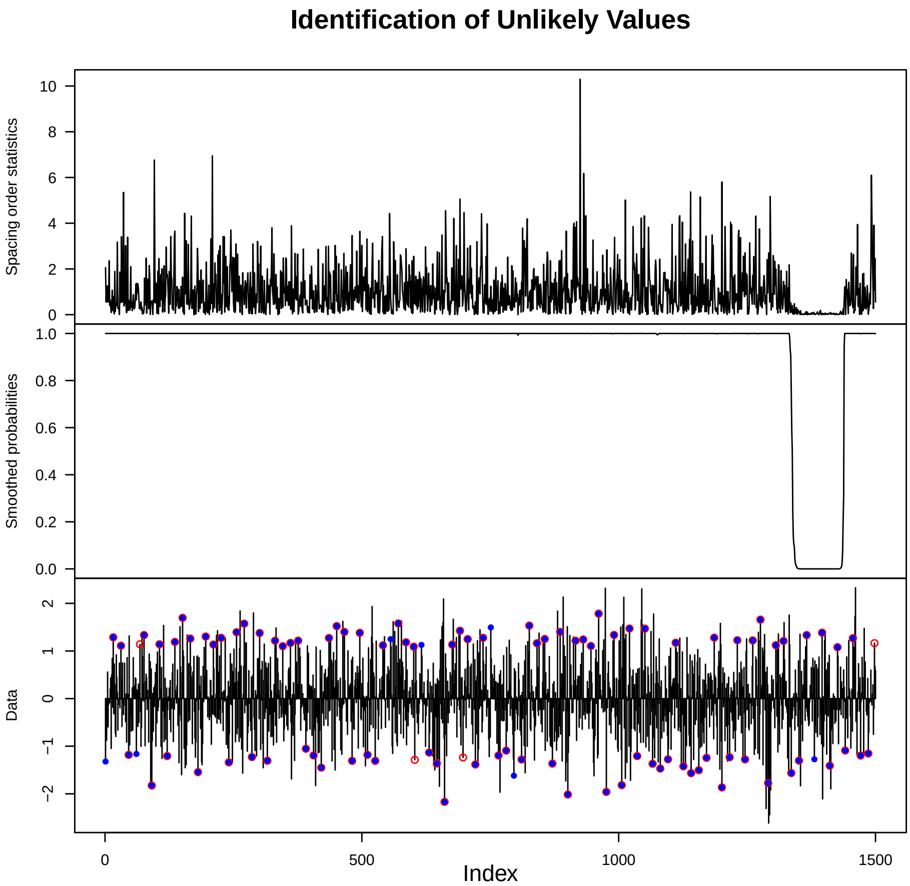

Figure 2: Estimation of a contaminated GARCH time series with the auto-WTLE estimator. The GARCH(1,1) series is of length 1500 with parameters $\omega = \texttt{0.1}$, $\alpha = \texttt{0.2}$, and $\beta = \texttt{0.6}$ with 15 equidistant outliers of scale $\mathrm{d} = \texttt{5}$. The upper figure plots is the spacing order statistics. The middle figure displays the smoothed probabilites after applying the Mone-Carlo switching model filter on the spacings. The lower figure shows the time series. The empty circles are the points for which their likely values have been trimmed in the WTLE and the full cirlces are the exact outliers that were added to the time series.

4.3 Simulation study

All models considered within this study were implemented in the `R` statistical programming language (R Core Team). The computation times reported offer only an indication of performance and may change with the platform. Regardless, the purpose of this work was not to obtain the most efficient implementation. All models were implemented in `R`, except for the computation of the likelihood function, which was implemented in `C`.

In this section, the GARCH auto-WTLE is compared to the QML, the GARCH M-estimators (M1, M2), with their bounded versions (BM1, BM2) as 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, the robust parameters are set to the values recommended by the estimator authors. However, for the REC estimator, stronger threshold statistic ($c = 4$) was used, than those recommended by Charles and Darne (2005). Indeed, it was noticed that for large outliers, it is crucial to use a low threshold. Otherwise, the unfiltered outliers will lead to poor convergence rates for the optimization routines. The trimming parameter for the GARCH WTLE was automatically defined, as described in Section 3.

Since all of the methods compared are based on the MLE, the deviation of the estimates of the robust models with the contaminated series, from the maximum likelihood estimates of the respective uncontaminated series, is reported. The deviations, are defined as $\hat{\theta}^{(y)} - \hat{\theta}^{(x)}$, where $\hat{\theta}^{(y)}$ are the fitted parameters of the contaminated series y_t, and $\hat{\theta}^{(x)}$ are the maximum likelihood estimates of the uncontaminated series, x_t.

The mean absolute deviation (MAD) of the fitted parameters is defined as

$$\widehat{\text{MAD}} = \frac{1}{N} \sum_{i=1}^{N} |\hat{\theta}_i^{(y)} - \hat{\theta}_i^{(x)}|,$$

where N is the number of Monte-Carlo runs. Similarly, the mean square deviation (MSD) of the estimates of y_t with the maximum likelihood estimates of the uncontaminated series x_t, is expressed as

$$\widehat{\text{MSD}} = \frac{1}{N} \sum_{i=1}^{N} \left[\hat{\theta}_i^{(y)} - \hat{\theta}_i^{(x)}\right]^2.$$

For all models, the same initial values are used for the conditional variance. Indeed, due to the recursive nature of the GARCH model, initial values must be provided for both ε_0^2 and σ_0^2. All models use the unconditional variance of the uncontaminated series calculated from the simulation parameters. With normal innovations, the uncontaminated series becomes,

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_i^q \alpha_i - \sum_i^q \beta_i}.$$

Note that the same starting values are used in the routine of each of the different estimators. Starting values that were different from the parameter values used to generate the data set were explicitly chosen in order to assess the ability of the estimator to converge to a solution.

	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.
REC	0.018 (5.10E-04)	0.244 (6.09E-02)	0.145 (2.40E-02)	100%
M1	0.011 (2.32E-04)	0.034 (2.13E-03)	0.026 (1.26E-03)	100%
BM1	0.017 (5.62E-04)	0.054 (4.50E-03)	0.042 (3.10E-03)	93%
M2	0.020 (7.78E-04)	0.058 (5.40E-03)	0.045 (3.65E-03)	99%
BM2	0.017 (5.45E-04)	0.116 (1.74E-02)	0.046 (3.58E-03)	86%
WTL	0.000 (1.45E-07)	0.000 (1.48E-06)	0.000 (9.49E-07)	100%

(a) $\omega = 0.1$, $\alpha = 0.5$ and $\beta = 0.4$

	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.
REC	0.047 (6.65E-03)	0.049 (2.84E-03)	0.075 (1.29E-02)	99%
M1	0.026 (1.64E-03)	0.014 (3.47E-04)	0.035 (2.91E-03)	87%
BM1	0.044 (5.24E-03)	0.026 (1.23E-03)	0.059 (8.71E-03)	95%
M2	0.045 (5.62E-03)	0.024 (9.82E-04)	0.060 (8.92E-03)	91%
BM2	0.044 (5.44E-03)	0.046 (3.22E-03)	0.063 (9.46E-03)	90%
WTL	0.000 (6.10E-07)	0.000 (1.40E-07)	0.000 (1.03E-06)	100%

(b) $\omega = 0.1$, $\alpha = 0.1$ and $\beta = 0.8$

Table 1: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) to assess the estimators' bias to the MLE. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

Table 1 lists the MADs and MSDs of the models with uncontaminated series. This is in order to show how the estimators are biased from the classical MLE. It is clear that the auto-WTLE deviates very little from the MLE estimates and has the smallest MSD compared to the other models. The auto-WTLE thus has almost no bias to the MLE with uncontaminated series. The table also shows the convergence success rate of the optimization routine. The convergence success rate corresponds to the percentage of Monte-Carlo iterations for which the optimization routine did converge to a solution, according to the convergence criterion of the optimization routine. Note these values are specific to the optimization algorithm used (`nlminb()` function in `R`) and might be different for other algorithms.

To make a second comparison, some 1000 sequences were generated, each of 1500 GARCH(1,1) simulated sample variates as described in Eq. (3), with 1%, 5%, and 10% outliers. The contaminated time series, y_t, were constructed from the uncontaminated series, x_t, as $y_t = x_t$ for $t \neq i$ plus outliers $y_i = d\,\sigma_i$ at time index i and scale d. A range of outlier scales, $d \in \{2, 4, 6, 10\}$, was used in order to study how the methods perform with slightly and greatly abnormal points. The outliers were taken from the truncated Poisson distribution with a truncation of 10.

Two sets of parameters were used for the GARCH(1,1) model and the starting values in the optimization routines were explicitly set to be different from the optimal values. Tables 2 and 3 displays results from the simulation using GARCH(1,1) with parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. Although these values are not typically encountered when dealing with financial returns (i.e., a large β and small α with a persistence, $\alpha + \beta$, close to 1), they were chosen in order to compare the results with those of Muler and Yohai (2008). For Tables 4 and 5, more realistic parameter values of $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$, were used. As noted in Charles and Darne (2005), the bounded M-estimators (BM1 and BM2) have a smaller bias than the unbounded M-estimators (M1 and M2), but are subject to a lower convergence success rate. Moreover, M2 and BM2 produce better estimates than do their less-robust counterparts, M1 and BM1. Although the estimates of the REC estimator are similar to those of the other estimators, the REC computation time was much larger than those of the others. Overall, the WTLE yields estimates with the smallest MSE and RMSE, and yet it incurs only a small computational cost.

Besides the MADs and MSDs shown in the tables, also consider the empirical distributions of the deviation of the fitted parameters of the robust models from the fitted parameters obtained with the MLE of the uncontaminated series. Figures 3 and 4 display results for the REC, BM2 and auto-WTL estimators. Only these three estimators are compared because the parameter distributions of the M2 model are too wide to be included within the same graphics. It is clear that the deviations of the auto-WTLE estimates have the smallest variance, and are closer to zero for the GARCH parameter values.

5 Conclusion

The robust estimation of parameters is essential in practice. This is especially true for financial applications where outliers are common in the data. There are many possible origins for these outliers. An outlier may arise from a recording error or from an abrupt regime change due to either a political decision or a rumor. The classical econometric models, such as the GARCH model, can be very sensitive to abnormal points. However, such models are often the foundation of the risk measures that are used by large financial companies to guide their investment strategies. Consequently, there is much to be gained from improving the robustness of econometric models.

In this paper, it was shown how to transform the maximum likelihood estimator into a robust estimator. The contribution of the present work, is to construct an automatic selection process for the WTLE parameters. The proposed fully automatic method for selecting the trimming parameter and the weights in the WTLE, obviates the tedious fine tuning process required by other robust models.

The auto-WTLE was successfully applied to GARCH modeling. It was shown, through an extensive simulation study, that the auto-WTLE provides robust and reliable estimates at only a small computational cost. Note that only the simple GARCH(1,1) model was considered.

1% with d=2.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.005 (3.44E-05)	0.009 (1.92E-04)	0.009 (1.45E-04)	100%	0.01
REC	0.017 (4.88E-04)	0.249 (6.36E-02)	0.148 (2.47E-02)	100%	2.19
M1	0.012 (2.71E-04)	0.036 (2.30E-03)	0.027 (1.26E-03)	100%	0.04
BM1	0.020 (7.04E-04)	0.058 (5.28E-03)	0.046 (3.44E-03)	92%	0.06
M2	0.023 (9.18E-04)	0.063 (6.04E-03)	0.047 (3.95E-03)	99%	0.04
BM2	0.020 (6.51E-04)	0.115 (1.72E-02)	0.047 (3.50E-03)	87%	0.07
WTL	0.005 (3.37E-05)	0.009 (1.25E-04)	0.009 (1.33E-04)	100%	0.02

1% with d=4.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.030 (1.12E-03)	0.056 (5.63E-03)	0.039 (2.28E-03)	100%	0.01
REC	0.017 (5.45E-04)	0.259 (6.84E-02)	0.150 (2.55E-02)	100%	2.45
M1	0.017 (5.12E-04)	0.038 (2.38E-03)	0.040 (2.67E-03)	100%	0.04
BM1	0.019 (6.34E-04)	0.050 (3.90E-03)	0.046 (3.51E-03)	91%	0.06
M2	0.026 (1.21E-03)	0.060 (5.63E-03)	0.061 (6.48E-03)	99%	0.05
BM2	0.019 (6.22E-04)	0.099 (1.33E-02)	0.048 (3.84E-03)	87%	0.07
WTL	0.006 (1.25E-04)	0.012 (2.81E-04)	0.011 (2.69E-04)	100%	0.03

5% with d=2.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.026 (7.54E-04)	0.022 (1.50E-03)	0.022 (9.53E-04)	100%	0.01
REC	0.025 (1.05E-03)	0.268 (7.35E-02)	0.136 (2.20E-02)	100%	2.20
M1	0.034 (1.51E-03)	0.044 (3.62E-03)	0.035 (2.63E-03)	100%	0.04
BM1	0.050 (3.35E-03)	0.080 (9.73E-03)	0.051 (5.29E-03)	93%	0.05
M2	0.049 (3.42E-03)	0.086 (1.17E-02)	0.053 (5.58E-03)	98%	0.04
BM2	0.048 (3.18E-03)	0.114 (1.75E-02)	0.056 (5.24E-03)	87%	0.06
WTL	0.022 (5.99E-04)	0.019 (6.31E-04)	0.020 (6.62E-04)	100%	0.04

5% with d=4.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.240 (5.98E-02)	0.191 (4.91E-02)	0.151 (2.75E-02)	92%	0.01
REC	0.049 (5.00E-03)	0.321 (1.05E-01)	0.148 (2.78E-02)	100%	3.18
M1	0.063 (5.85E-03)	0.158 (2.98E-02)	0.109 (1.89E-02)	100%	0.05
BM1	0.023 (9.20E-04)	0.104 (1.41E-02)	0.059 (5.94E-03)	88%	0.06
M2	0.074 (8.15E-03)	0.174 (4.07E-02)	0.152 (3.44E-02)	99%	0.06
BM2	0.020 (6.87E-04)	0.060 (5.58E-03)	0.055 (4.84E-03)	84%	0.07
WTL	0.007 (9.33E-05)	0.015 (4.48E-04)	0.017 (5.52E-04)	100%	0.02

10% with d=2.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.046 (2.36E-03)	0.029 (2.68E-03)	0.033 (2.22E-03)	99%	0.01
REC	0.048 (3.31E-03)	0.285 (8.34E-02)	0.124 (2.00E-02)	100%	2.22
M1	0.061 (4.60E-03)	0.058 (7.78E-03)	0.046 (4.65E-03)	100%	0.04
BM1	0.081 (8.00E-03)	0.103 (1.55E-02)	0.060 (7.54E-03)	95%	0.06
M2	0.103 (1.32E-02)	0.129 (2.39E-02)	0.061 (7.67E-03)	99%	0.04
BM2	0.081 (7.93E-03)	0.121 (2.07E-02)	0.064 (7.59E-03)	85%	0.07
WTL	0.041 (2.03E-03)	0.026 (1.32E-03)	0.030 (1.67E-03)	100%	0.03

10% with d=4.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.470 (2.26E-01)	0.224 (6.50E-02)	0.224 (5.62E-02)	82%	0.01
REC	0.116 (2.45E-02)	0.372 (1.40E-01)	0.170 (4.30E-02)	99%	3.46
M1	0.079 (1.35E-02)	0.288 (9.16E-02)	0.152 (3.38E-02)	99%	0.06
BM1	0.022 (8.85E-04)	0.151 (2.65E-02)	0.065 (6.72E-03)	90%	0.06
M2	0.073 (9.57E-03)	0.307 (1.09E-01)	0.199 (5.26E-02)	96%	0.07
BM2	0.021 (7.57E-04)	0.055 (4.64E-03)	0.054 (5.02E-03)	83%	0.07
WTL	0.007 (1.33E-04)	0.018 (5.76E-04)	0.020 (7.39E-04)	100%	0.02

Table 2: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{2, 4\}$ and parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

1% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.076 (7.07E-03)	0.154 (3.48E-02)	0.082 (9.72E-03)	98%	0.01
REC	0.018 (5.21E-04)	0.251 (6.43E-02)	0.150 (2.52E-02)	100%	2.51
M1	0.031 (1.39E-03)	0.038 (2.32E-03)	0.071 (7.31E-03)	100%	0.04
BM1	0.019 (6.34E-04)	0.050 (3.90E-03)	0.046 (3.51E-03)	91%	0.06
M2	0.039 (2.49E-03)	0.063 (6.29E-03)	0.094 (1.39E-02)	100%	0.05
BM2	0.019 (6.21E-04)	0.099 (1.33E-02)	0.048 (3.84E-03)	87%	0.07
WTL	0.002 (9.09E-06)	0.006 (6.79E-05)	0.006 (5.87E-05)	100%	0.02

1% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.251 (7.95E-02)	0.272 (9.54E-02)	0.138 (2.88E-02)	51%	0.01
REC	0.018 (5.47E-04)	0.251 (6.44E-02)	0.151 (2.57E-02)	100%	2.52
M1	0.062 (4.96E-03)	0.049 (3.72E-03)	0.146 (2.54E-02)	100%	0.04
BM1	0.019 (6.34E-04)	0.050 (3.90E-03)	0.046 (3.51E-03)	91%	0.06
M2	0.069 (6.87E-03)	0.077 (8.88E-03)	0.170 (3.61E-02)	99%	0.05
BM2	0.019 (6.21E-04)	0.099 (1.33E-02)	0.048 (3.84E-03)	87%	0.07
WTL	0.002 (8.93E-06)	0.006 (6.77E-05)	0.006 (5.82E-05)	100%	0.02

5% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.710 (5.25E-01)	0.334 (1.29E-01)	0.254 (7.39E-02)	29%	0.01
REC	0.019 (6.49E-04)	0.290 (8.59E-02)	0.175 (3.47E-02)	100%	3.51
M1	0.150 (2.64E-02)	0.137 (2.88E-02)	0.270 (8.48E-02)	96%	0.05
BM1	0.023 (9.19E-04)	0.104 (1.41E-02)	0.059 (5.94E-03)	88%	0.06
M2	0.155 (2.90E-02)	0.142 (3.52E-02)	0.310 (1.06E-01)	90%	0.06
BM2	0.020 (6.87E-04)	0.060 (5.58E-03)	0.055 (4.85E-03)	84%	0.07
WTL	0.004 (3.39E-05)	0.012 (2.88E-04)	0.013 (2.85E-04)	100%	0.02

5% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	1.695 (4.41E+00)	0.395 (1.63E-01)	0.346 (1.47E-01)	1%	0.02
REC	0.023 (9.28E-04)	0.297 (9.07E-02)	0.183 (3.83E-02)	100%	3.52
M1	0.210 (4.95E-02)	0.091 (1.71E-02)	0.386 (1.52E-01)	83%	0.05
BM1	0.023 (9.19E-04)	0.104 (1.41E-02)	0.059 (5.94E-03)	88%	0.06
M2	0.191 (4.37E-02)	0.111 (2.16E-02)	0.386 (1.52E-01)	74%	0.06
BM2	0.020 (6.87E-04)	0.060 (5.58E-03)	0.055 (4.85E-03)	84%	0.07
WTL	0.004 (3.32E-05)	0.012 (2.61E-04)	0.012 (2.52E-04)	100%	0.02

10% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	1.342 (1.83E+00)	0.335 (1.30E-01)	0.314 (1.09E-01)	14%	0.01
REC	0.026 (1.19E-03)	0.328 (1.10E-01)	0.209 (4.95E-02)	100%	4.19
M1	0.125 (2.46E-02)	0.313 (1.19E-01)	0.302 (1.01E-01)	94%	0.07
BM1	0.022 (8.83E-04)	0.151 (2.65E-02)	0.065 (6.71E-03)	90%	0.07
M2	0.125 (2.52E-02)	0.294 (1.16E-01)	0.311 (1.08E-01)	87%	0.08
BM2	0.020 (7.51E-04)	0.055 (4.62E-03)	0.054 (4.98E-03)	83%	0.07
WTL	0.005 (5.25E-05)	0.015 (3.59E-04)	0.016 (3.99E-04)	100%	0.02

10% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.579 (8.78E-01)	0.474 (2.28E-01)	0.503 (2.59E-01)	11%	0.03
REC	0.059 (1.34E-02)	0.338 (1.18E-01)	0.208 (5.29E-02)	98%	4.46
M1	0.158 (3.66E-02)	0.263 (1.03E-01)	0.362 (1.39E-01)	84%	0.08
BM1	0.022 (8.83E-04)	0.151 (2.65E-02)	0.065 (6.71E-03)	90%	0.06
M2	0.198 (6.53E-02)	0.159 (4.92E-02)	0.366 (1.42E-01)	79%	0.07
BM2	0.020 (7.51E-04)	0.055 (4.62E-03)	0.054 (4.98E-03)	83%	0.07
WTL	0.005 (5.08E-05)	0.015 (3.53E-04)	0.016 (3.89E-04)	100%	0.02

Table 3: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{6, 10\}$ and parameters $\omega = 0.1$, $\alpha = 0.5$, and $\beta = 0.4$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

	1% with d=2.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.009 (1.44E-04)	0.005 (4.20E-05)	0.012 (2.42E-04)	100%	0.02
REC	0.051 (8.53E-03)	0.051 (3.01E-03)	0.077 (1.43E-02)	98%	1.29
M1	0.026 (1.51E-03)	0.014 (3.42E-04)	0.034 (2.37E-03)	90%	0.05
BM1	0.051 (7.71E-03)	0.029 (1.46E-03)	0.066 (1.08E-02)	92%	0.07
M2	0.055 (8.83E-03)	0.027 (1.24E-03)	0.068 (1.17E-02)	88%	0.06
BM2	0.053 (8.82E-03)	0.046 (3.45E-03)	0.071 (1.30E-02)	91%	0.07
WTL	0.008 (1.43E-04)	0.005 (4.19E-05)	0.012 (2.42E-04)	100%	0.02

	1% with d=4.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.078 (1.24E-02)	0.022 (7.67E-04)	0.063 (9.87E-03)	100%	0.01
REC	0.047 (6.39E-03)	0.052 (3.16E-03)	0.076 (1.25E-02)	99%	1.69
M1	0.049 (6.32E-03)	0.029 (1.16E-03)	0.061 (8.39E-03)	90%	0.05
BM1	0.049 (7.55E-03)	0.025 (1.13E-03)	0.066 (1.18E-02)	93%	0.06
M2	0.071 (1.44E-02)	0.035 (1.73E-03)	0.091 (2.00E-02)	89%	0.06
BM2	0.050 (8.01E-03)	0.044 (3.14E-03)	0.071 (1.34E-02)	91%	0.07
WTL	0.008 (4.17E-04)	0.005 (7.98E-05)	0.012 (7.47E-04)	100%	0.02

	5% with d=2.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.032 (1.97E-03)	0.015 (3.01E-04)	0.028 (1.44E-03)	100%	0.02
REC	0.111 (3.30E-02)	0.059 (4.01E-03)	0.125 (3.35E-02)	96%	0.94
M1	0.053 (5.94E-03)	0.017 (4.88E-04)	0.045 (4.28E-03)	94%	0.05
BM1	0.104 (3.07E-02)	0.027 (1.27E-03)	0.087 (2.08E-02)	94%	0.07
M2	0.104 (2.92E-02)	0.025 (1.23E-03)	0.083 (1.92E-02)	88%	0.06
BM2	0.105 (3.18E-02)	0.034 (2.10E-03)	0.093 (2.30E-02)	92%	0.07
WTL	0.029 (1.79E-03)	0.013 (2.42E-04)	0.027 (1.37E-03)	100%	0.05

	5% with d=4.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.427 (4.21E-01)	0.085 (8.18E-03)	0.273 (1.12E-01)	26%	0.02
REC	0.053 (7.20E-03)	0.054 (3.41E-03)	0.085 (1.49E-02)	99%	2.95
M1	0.064 (1.03E-02)	0.072 (5.68E-03)	0.105 (1.73E-02)	90%	0.07
BM1	0.051 (7.45E-03)	0.026 (1.09E-03)	0.068 (1.19E-02)	94%	0.07
M2	0.087 (2.09E-02)	0.071 (5.60E-03)	0.123 (2.81E-02)	89%	0.08
BM2	0.050 (7.46E-03)	0.036 (2.23E-03)	0.070 (1.30E-02)	93%	0.07
WTL	0.017 (1.78E-03)	0.011 (3.10E-04)	0.026 (3.07E-03)	100%	0.02

	10% with d=2.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.031 (3.62E-03)	0.030 (1.08E-03)	0.051 (4.45E-03)	100%	0.02
REC	0.117 (4.62E-02)	0.071 (5.60E-03)	0.154 (4.14E-02)	91%	0.82
M1	0.061 (1.44E-02)	0.030 (1.21E-03)	0.071 (1.08E-02)	84%	0.06
BM1	0.089 (3.28E-02)	0.034 (1.69E-03)	0.098 (1.94E-02)	95%	0.08
M2	0.113 (3.84E-02)	0.041 (2.79E-03)	0.093 (1.92E-02)	82%	0.06
BM2	0.107 (4.73E-02)	0.035 (1.89E-03)	0.115 (2.72E-02)	91%	0.09
WTL	0.033 (4.78E-03)	0.026 (8.39E-04)	0.048 (4.82E-03)	100%	0.04

	10% with d=4.0 $\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.037 (4.67E-03)	0.105 (1.15E-02)	0.158 (2.73E-02)	11%	0.03
REC	0.059 (1.02E-02)	0.050 (3.10E-03)	0.093 (2.07E-02)	99%	3.74
M1	0.054 (6.13E-03)	0.076 (6.25E-03)	0.096 (1.36E-02)	94%	0.07
BM1	0.042 (4.06E-03)	0.026 (1.11E-03)	0.057 (7.22E-03)	94%	0.07
M2	0.068 (1.16E-02)	0.075 (6.20E-03)	0.104 (1.87E-02)	89%	0.08
BM2	0.042 (4.65E-03)	0.034 (1.91E-03)	0.062 (9.08E-03)	91%	0.07
WTL	0.020 (1.46E-03)	0.014 (3.87E-04)	0.032 (2.83E-03)	100%	0.02

Table 4: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{2, 4\}$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

1% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.332 (1.66E-01)	0.057 (5.15E-03)	0.224 (8.99E-02)	87%	0.02
REC	0.049 (6.94E-03)	0.052 (3.11E-03)	0.077 (1.34E-02)	99%	1.67
M1	0.114 (3.15E-02)	0.052 (3.23E-03)	0.128 (3.54E-02)	88%	0.06
BM1	0.049 (7.55E-03)	0.025 (1.13E-03)	0.066 (1.18E-02)	93%	0.06
M2	0.133 (4.16E-02)	0.057 (4.04E-03)	0.154 (5.01E-02)	84%	0.08
BM2	0.050 (8.01E-03)	0.044 (3.14E-03)	0.071 (1.34E-02)	91%	0.07
WTL	0.004 (6.18E-05)	0.003 (1.37E-05)	0.006 (1.09E-04)	100%	0.02

1% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.834 (8.11E-01)	0.292 (1.21E-01)	0.466 (2.60E-01)	16%	0.01
REC	0.053 (8.83E-03)	0.052 (3.18E-03)	0.082 (1.63E-02)	99%	1.65
M1	0.198 (8.19E-02)	0.082 (7.55E-03)	0.211 (8.45E-02)	69%	0.08
BM1	0.049 (7.55E-03)	0.025 (1.13E-03)	0.066 (1.18E-02)	93%	0.07
M2	0.225 (9.67E-02)	0.087 (8.36E-03)	0.233 (9.83E-02)	63%	0.10
BM2	0.050 (8.01E-03)	0.044 (3.14E-03)	0.071 (1.34E-02)	91%	0.07
WTL	0.004 (6.24E-05)	0.003 (1.41E-05)	0.006 (1.12E-04)	100%	0.02

5% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML				0%	
REC	0.068 (1.40E-02)	0.057 (3.72E-03)	0.106 (2.77E-02)	98%	2.89
M1	0.099 (2.88E-02)	0.091 (8.70E-03)	0.149 (3.65E-02)	82%	0.09
BM1	0.051 (7.45E-03)	0.026 (1.09E-03)	0.068 (1.19E-02)	94%	0.07
M2	0.133 (4.71E-02)	0.089 (8.47E-03)	0.172 (5.19E-02)	77%	0.10
BM2	0.050 (7.46E-03)	0.036 (2.23E-03)	0.070 (1.30E-02)	93%	0.07
WTL	0.009 (2.18E-04)	0.006 (5.60E-05)	0.014 (4.24E-04)	100%	0.02

5% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML				0%	
REC	0.067 (1.29E-02)	0.058 (3.85E-03)	0.105 (2.58E-02)	98%	2.96
M1	0.154 (5.95E-02)	0.098 (1.01E-02)	0.184 (5.74E-02)	64%	0.10
BM1	0.051 (7.45E-03)	0.026 (1.09E-03)	0.068 (1.19E-02)	94%	0.07
M2	0.171 (6.77E-02)	0.098 (1.02E-02)	0.198 (6.72E-02)	64%	0.11
BM2	0.050 (7.46E-03)	0.036 (2.23E-03)	0.070 (1.30E-02)	93%	0.07
WTL	0.009 (2.11E-04)	0.005 (5.22E-05)	0.013 (4.03E-04)	100%	0.02

10% with d=6.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML	0.059 (4.35E-03)	0.109 (1.21E-02)	0.150 (2.29E-02)	0%	0.03
REC	0.072 (1.44E-02)	0.058 (4.00E-03)	0.114 (2.99E-02)	96%	3.64
M1	0.072 (1.40E-02)	0.091 (8.83E-03)	0.126 (2.42E-02)	89%	0.08
BM1	0.042 (4.06E-03)	0.026 (1.11E-03)	0.057 (7.22E-03)	94%	0.07
M2	0.092 (2.40E-02)	0.091 (8.84E-03)	0.137 (3.25E-02)	79%	0.10
BM2	0.042 (4.65E-03)	0.034 (1.91E-03)	0.062 (9.08E-03)	91%	0.07
WTL	0.013 (4.20E-04)	0.008 (1.26E-04)	0.021 (8.73E-04)	100%	0.02

10% with d=10.0	$\tilde{\omega}$	$\tilde{\alpha}$	$\tilde{\beta}$	conv.	time [s]
QML				0%	
REC	0.071 (1.35E-02)	0.062 (4.49E-03)	0.114 (2.87E-02)	93%	3.63
M1	0.114 (3.80E-02)	0.099 (1.03E-02)	0.162 (4.38E-02)	70%	0.10
BM1	0.042 (4.06E-03)	0.026 (1.11E-03)	0.057 (7.22E-03)	94%	0.07
M2	0.139 (5.02E-02)	0.098 (1.02E-02)	0.174 (5.38E-02)	63%	0.11
BM2	0.042 (4.65E-03)	0.034 (1.91E-03)	0.062 (9.08E-03)	91%	0.07
WTL	0.012 (3.66E-04)	0.008 (9.84E-05)	0.018 (7.14E-04)	100%	0.02

Table 5: Mean square deviation and relative mean square deviation for the simple GARCH(1,1) with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d \in \{6, 10\}$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 long with a 500 burn-in sequence. The number of Monte-Carlo replications is 1000. We also report the percentage count of convergence of the optimization routines and the elapsed computation time in seconds.

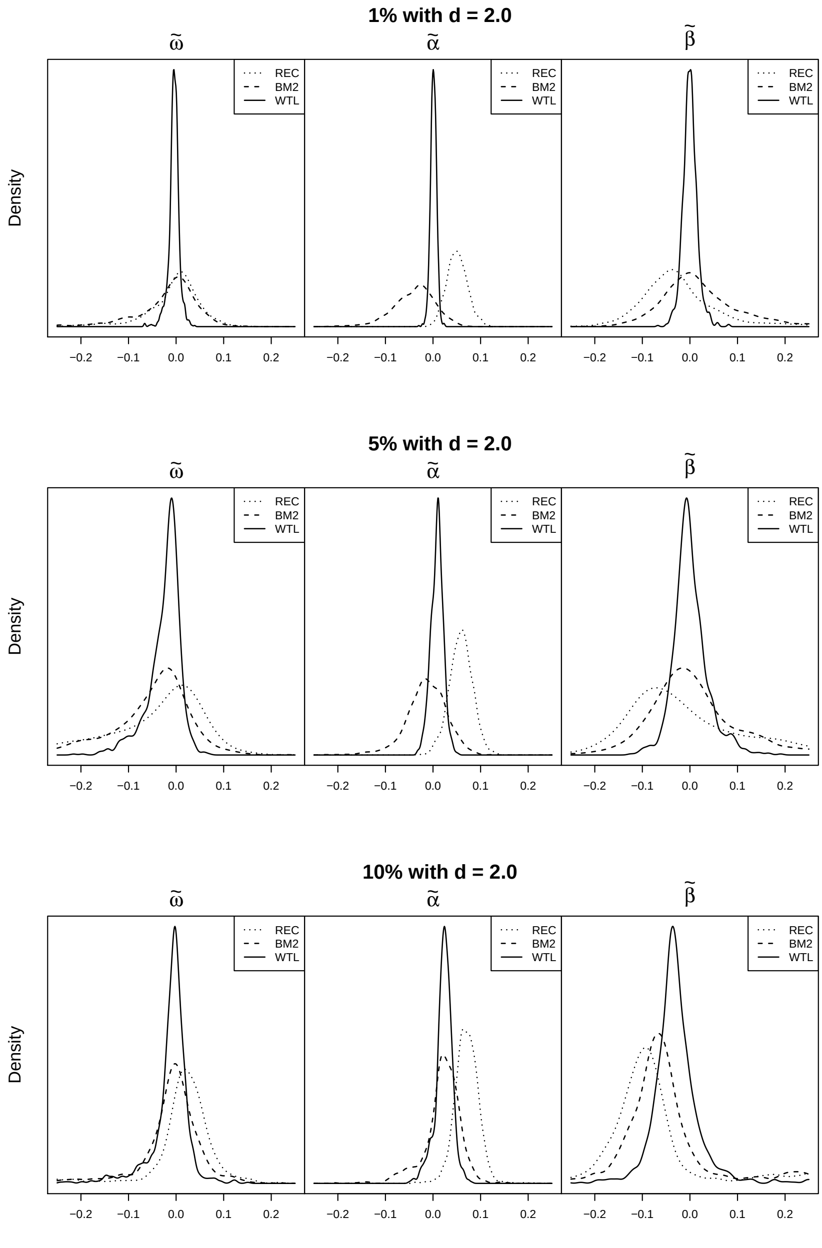

Figure 3: Kernel density approximation of the deviations for the simple GARCH(1,1) for REC, BME and WTLE with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d = 2$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 with 500 burn-in sequence. The number of Monte-Carlo replications is 1000.

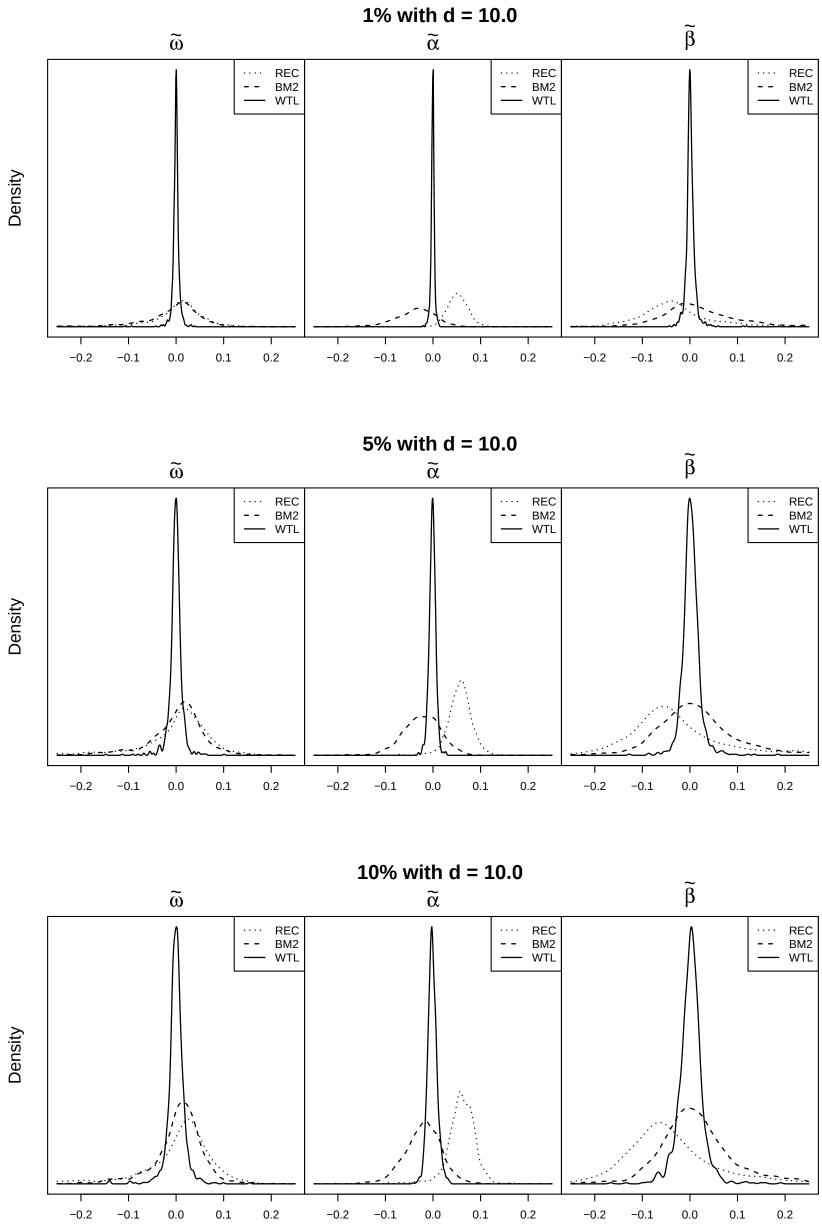

Figure 4: Kernel density approximation of the deviations for the simple GARCH(1,1) for REC, BME and WTLE with $\{1\%, 5\%, 10\%\}$ of outliers, $y_i = d\,\sigma_i$, with scale $d = 10$ and parameters $\omega = 0.1$, $\alpha = 0.1$, and $\beta = 0.8$. The length of the simulated series is 1500 with 500 burn-in sequence. The number of Monte-Carlo replications is 1000.

However, the WTLE can be used with any model for which there exists a likelihood estimator.

The automatic calculation procedure for the weights and trimming parameters in the WTLE could be used in other applications. For example, the size of the trimming parameters could be used as stability measure, which would indicate how many of the data points could be accurately represented by the distribution model used. When the trimming parameters significantly increase in a time horizon, this is a clear indication that the underlying model is inadequate and that a new regime might be in effect. Such changes could be represented in terms of a stability measure.

Acknowledgments

This work is part of the PhD thesis of Yohan Chalabi. He acknowledges financial support from Finance Online GmbH and from the Swiss Federal Institute of Technology Zurich.

References

T. Bednarski and B. Clarke. Trimmed likelihood estimation of location and scale of the normal distribution. *Australian Journal of Statistics*, 35(2):141–153, 1993.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, Apr. 1986.

T. Bollerslev. Glossary to ARCH (GARCH). Feb. 2009.

C. Brooks, S. Burke, and G. Persand. Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting*, 17(1):45–56, 2001.

A. Charles and O. Darne. Outliers and GARCH models in financial data. *Economics Letters*, 86(3):347–352, Mar. 2005.

R. C. H. Cheng and N. A. K. Amin. Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45 (3):394–403, 1983.

P. Cizek. Greneral trimmed estimation: robust approach to nonlinear and limited dependent variable models. *Econometric Theory*, 24(06):1500–1529, 2008.

H. David and H. Nagaraja. Order statistics. *NJ: John Wiley & Sons*, 7:159–61, 2003.

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

R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Lond. Phil. Trans. (A)*, 223:1–33, 1922.

A. Hadi and A. Luceño. Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. *Computational Statistics & Data Analysis*, 25(3):251–272, 1997.

L. Hotta and R. Tsay. Outliers in GARCH processes, 1998.

C.-J. Kim. Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1–2): 1–22, 1994.

C. Kuan. Lecture on the Markov switching model. Institute of Economics Academia Sinica, 2002.

M. Markatou. Mixture models, robustness, and the weighted likelihood methodology. *Biometrics*, 56(2):483–486, 2000.

B. Mendes. Assessing the bias of maximum likelihood estimates of contaminated GARCH models. *Journal of Statistical Computation and Simulation*, 67(4):359–376, 2000.

N. Muler and V. Yohai. Robust estimates for GARCH models. *Journal of Statistical Planning and Inference*, 138(10):2918–2940, 2008.

C. Müller and N. Neykov. Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference*, 116(2): 503–519, 2003.

N. Neykov and C. Müller. Breakdown point and computation of trimmed likelihood estimators in generalized linear models. *Developments in robust statistics. Physica-Verlag, Heidelberg*, pages 277–286, 2003.

N. Neykov and P. Neytchev. A robust alternative of the maximum likelihood estimators. *COMPSTAT 1990 - Short Communications*, pages 99–100, 1990.

N. Neykov, P. Filzmoser, R. Dimova, and P. Neytchev. Robust fitting of mixtures using the trimmed likelihood estimator. *Computational Statistics & Data Analysis*, 52(1):299–308, 2007.

R. Pyke. Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 395–449, 1965.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

B. Ranneby. The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 11(2):93–112, 1984.

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

D. Vandev and N. Neykov. About regression estimators with high breakdown point. *Statistics*, 32(2):111–129, 1998.

About the Authors

Diethelm Würtz is Private Lecturer at the "Institute for Theoretical Physics" at the Swiss Federal Institute of Technology (ETH) in Zurich. His research interests are in the field of risk management and stability analysis of financial markets. He teaches computational science and financial engineering. He is senior partner of the ETH spin-off company "Finance Online" and president of the "Rmetrics Association in Zurich".

Yohan Chalabi has a master in Physics from the Swiss Federal Institute of Technology in Lausanne. He is a PhD student in the Econophysics group at ETH Zurich at the Institute for Theoretical Physics. Yohan is a maintainer of the Rmetrics packages and the R/Rmetrics software environment.

Acknowledgement

The work presented in this article was partly supported by grants given by ETH Zurich and Rmetrics Association Zurich.