Biometrika (2018), **103**, 1, *pp*. 1–8
Printed in Great Britain Advance Access publication on 31 July 2018

Bootstrapping M-estimators in GARCH models

BY K. MUKHERJEE
Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, U.K.
k.mukherjee@lancaster.ac.uk

SUMMARY

We consider the weighted bootstrap approximation of the distribution of a class of M-estimators of the GARCH (p, q) parameters. We prove that the bootstrap distribution, given the data, is a consistent estimate in probability of the distribution of the M-estimator which is asymptotically normal. We propose an algorithm for the computation of M-estimates which at the same time is software-friendly to compute the bootstrap replicates from the given data. Our simulation study indicates superior coverage rates for various weighted bootstrap schemes compared with the rates based on the normal approximation and the existing bootstrap methods in the literature such as percentile t-subsampling schemes for the GARCH model. Since some familiar bootstrap schemes are special cases of the weighted bootstrap, this paper thus provides a unified theory and algorithm for bootstrapping in GARCH models.

Some key words: GARCH model; M-estimation; Weighted bootstrap.

1. INTRODUCTION

Consider the generalized autoregressive conditional heteroscedastic (GARCH) model of order (p, q) to analyze the volatility or the instantaneous variability of a financial time series $\{X_t; 1 \le t \le n\}$. Here the following representation of $\{X_t; t \in \mathcal{Z}\}$ is assumed:

$$X_t = \sigma_t \epsilon_t, \tag{1}$$

where $\{\epsilon_t; t \in \mathcal{Z}\}$ are unobservable independent and identically distributed errors with symmetric distribution around zero and

$$\sigma_t = \Big(\omega_0 + \sum_{i=1}^{p} \alpha_{0i} X_{t-i}^2 + \sum_{j=1}^{q} \beta_{0j} \sigma_{t-j}^2\Big)^{1/2}, \ t \in \mathcal{Z}, \tag{2}$$

with ω_0, α_{0i}, $\beta_{0j} > 0$, for all i, j. In this article, we show the asymptotic validity of a class of weighted bootstrap approximations of the distributions of M-estimators of the parameter

$$\theta_0 = (\omega_0, \alpha_{01}, \ldots, \alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})^{\mathrm{T}}$$

based on observations $\{X_t; 1 \le t \le n\}$. We propose an algorithm to compute M-estimators and their bootstrapped versions. Finally, we provide empirical evidence that the weighted bootstrap has better performance compared with those of the existing bootstrap schemes for GARCH models. Since our theoretical results on the asymptotic validity of weighted bootstrap estimators are proved under weak moment assumptions on errors, these are applicable to GARCH modeling of financial data where the quasi maximum likelihood estimator (QMLE) is routinely used for estimating parameters, even though higher moment assumptions for its asymptotics may not hold. Since some familiar bootstrap schemes are special cases of the weighted bootstrap, our result and algorithm provide a unified theory and computation of such schemes in GARCH models.

Mukherjee (2008) derived theoretical properties and demonstrated usefulness of M-estimators as a better alternative to the QMLE of the parameters of the GARCH model (1) and (2). See the references therein for other works in this area. Although M-estimators are asymptotically normal, their finite-sample distributions can be asymmetric. Consequently, confidence intervals based on the normality assumption do not have good coverage probabilities. Hence it is natural to explore the effectiveness of bootstrap procedures to approximate the finite-sample distributions of various M-estimators of the GARCH parameters.

For bootstrap-related work in such models, we mention Hidalgo and Zaffaroni (2007) who considered with-replacement bootstrap samples of standardized residuals and showed the first order consistency of some bootstrapped test statistics for checking the validity of the autoregressive conditional heteroscedastic (ARCH)(∞) model. Linton, Pan and Wang (2010) discussed a log-transformed model of squared observations of semi-strong GARCH model with some assumptions on the log-transformed error distribution, and investigated asymptotics of the least absolute deviation estimator. Hall and Yao (2003) and Linton, Pan and Wang (2010) also considered percentile t-subsampling bootstrap of the QMLE based on residuals. However, the weighted bootstrap is not subsumed by the residual bootstrap and other types of bootstrap discussed by these authors. Varga and Zempleni (2012) discussed the potential of the weighted bootstrap to approximate the distribution of the QMLE of the GARCH parameters and provided some empirical results for the ARCH(1) model that showed better performance of a particular weighted bootstrap, called Scheme M in this paper, than the residual bootstrap. Our empirical study shows better performance of the weighted bootstrap in comparison with existing bootstrap methods for the GARCH and related models in the literature for a wide spectrum of heavy-tailed as well as light-tailed error distributions.

Motivated by an algorithm proposed by Mak (1993) and applied later by Mak, Wong and Li (1997) in the context of ARCH type models for computing the QMLE, we propose a new algorithm to compute M-estimates in the GARCH model. We use a variant of this algorithm for computing bootstrap estimates as it avoids re-computation of some core quantities in new bootstrap samples and enables fast computation.

2. M-ESTIMATORS, WEIGHTED BOOTSTRAP AND ALGORITHM

Let $\psi : I\!R \to I\!R$ be an odd function, that is, $\psi(-x) = -\psi(x)$ for $x > 0$, which is differentiable at all but a finite number of points and let $H(x) := x\psi(x), x \in I\!R$. The function H is called the 'score function' for the M-estimation in the scale model.

Following Mukherjee (2008), we define M-estimators as follows. From Lemma 2.3 and Theorem 2.1 of Berkes et al. (2003), σ_t^2 of (2) has the unique almost sure representation $\sigma_t^2 = c_0 + \sum_{j=1}^{\infty} c_j X_{t-j}^2,\ t \in \mathcal{Z}$, where $\{c_j; j \geq 0\}$ are defined in (2.7)-(2.9) of Berkes et al. (2003). Let Θ be a compact subset of $(0,\infty)^{1+p} \times (0,1)^q$. A typical element in Θ is denoted by $\theta = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^{\mathrm{T}}$. Define the variance function on Θ by

$$v_t(\theta) = c_0(\theta) + \sum_{j=1}^{\infty} c_j(\theta) X_{t-j}^2, \quad \theta \in \Theta, t \in \mathcal{Z}, \tag{3}$$

where the coefficients $\{c_j(\theta); j \geq 0\}$ are given in Berkes et al. (2003, Section 3) with the property $c_j(\theta_0) = c_j, j \geq 0$, so that the variance functions satisfy $v_t(\theta_0) = \sigma_t^2, t \in \mathcal{Z}$ and

$$X_t = \{v_t(\theta_0)\}^{1/2} \epsilon_t, \ \ 1 \leq t \leq n. \tag{4}$$

Let $\{\hat{v}_t(\theta)\}$ be observable approximation of $\{v_t(\theta)\}$ of (3) defined by

$$\hat{v}_t(\theta) = c_0(\theta) + I(2 \leq t) \sum_{j=1}^{t-1} c_j(\theta) X_{t-j}^2, \ \theta \in \Theta, \ 1 \leq t \leq n.$$

Then an M-estimator $\hat{\theta}_n$ is defined as a solution of the system of equations

$$\sum_{t=1}^{n} [1 - H\{X_t/\hat{v}_t^{1/2}(\theta)\}]\{\dot{\hat{v}}_t(\theta)/\hat{v}_t(\theta)\} = 0, \tag{5}$$

where $\dot{\hat{v}}_t(\theta)$ is the first derivative of $\hat{v}_t(\theta)$. The QMLE corresponds to $\psi(x) = x$ while the least absolute deviation estimator corresponds to $\psi(x) = \text{sign}(x)$. The asymptotic distribution of $\hat{\theta}_n$ was derived in Mukherjee (2008) under various logarithmic moment conditions on the bounds of the increment of H (Conditions SM1-SM3 described in the Supplementary material) and the following identifiability condition on the existence of a unique number $c_H > 0$ such that

$$E\{H(\epsilon/c_H^{1/2})\} = 1. \tag{6}$$

Define the score function factor

$$\sigma^2(H) := 4\ \text{var}\{H(\epsilon/c_H^{1/2})\}/[E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}]^2,$$

the matrix $J := E\{\dot{v}_1(\theta_{0H})\dot{v}_1^{\mathrm{T}}(\theta_{0H})/v_1^2(\theta_{0H})\}$ and the transformed parameter

$$\theta_{0H} = (c_H\omega_0, c_H\alpha_{01}, \ldots, c_H\alpha_{0p}, \beta_{01}, \ldots, \beta_{0q})^{\mathrm{T}}.$$

Then the M-estimator $\hat{\theta}_n$ in (5) is asymptotically normal. As $n \to \infty$

$$n^{1/2}(\hat{\theta}_n - \theta_{0H}) \to N\{0, \sigma^2(H)J^{-1}\} \tag{7}$$

in distribution.

We now consider the weighted bootstrap formulation for M-estimators. Let $\{w_{nt}; 1 \leq t \leq n, n \geq 1\}$ be a triangular array of non-negative random variables with $E(w_{nt}) = 1$, such that for each $n \geq 1$, $\{w_{nt}; 1 \leq t \leq n\}$ are exchangeable and independent of the data $\{X_t; t \geq 1\}$ and errors $\{\epsilon_t; t \geq 1\}$.

Based on these weights, a bootstrap estimate $\hat{\theta}_{*n}$ is defined as a solution to the equation in θ

$$\sum_{t=1}^{n} w_{nt}[1 - H\{X_t/\hat{v}_t^{1/2}(\theta)\}]\{\dot{\hat{v}}_t(\theta)/\hat{v}_t(\theta)\} = 0.$$

Examples. From many different choices of bootstrap weights, we consider the following three schemes.

(i) Scheme M when $(w_{n1}, \ldots, w_{nn})$ have a multinomial $(n, 1/n, \ldots, 1/n)$ distribution. Here $\hat{\theta}_{*n}$ is the M-estimator computed from the units sampled which is essentially the classic paired bootstrap.

(ii) Scheme E when $w_{nt} = (nE_t)/\sum_{i=1}^{n} E_i$, where $\{E_t\}$ are independent and identically distributed exponential random variables with mean 1. Here $\hat{\theta}_{*n}$ is a weighted M-estimator with weights proportional to E_t, $1 \leq t \leq n$. This is an example of Bayesian bootstrap; see Praestgaard and Wellner (1993, p. 2058).

(iii) Scheme U when $w_{nt} = (nU_t)/\sum_{i=1}^{n} U_i$, where $\{U_t\}$ are independent and identically distributed uniform random variables from $U(1-a, 1+a)$ where $0 < a \leq 1$. Under Scheme U, $\hat{\theta}_{*n}$ is a weighted M-estimator with weights proportional to U_t, $1 \leq t \leq n$.

A number of other bootstrap methods in the literature are also special cases of the above bootstrap formulation. This general formulation of weighted bootstrap offers a unified way of studying several bootstrap schemes simultaneously.

In the following, E_* and var_* denote the expectation and variance with respect to the bootstrap distribution and all convergence statements hold as $n \to \infty$. We assume that the weights satisfy the following basic conditions (Conditions BW of Chatterjee and Bose 2005) where $\sigma_n^2 = \text{var}_*(w_{nt})$ and $k > 0$ is a constant:

$$E_*(w_{n1}) = 1,\ 0 < k < \sigma_n^2 = o(n),\ \text{corr}_*(w_{n1}, w_{n2}) = O(1/n), \tag{8}$$

and the standardized weights $\{W_{nt} = (w_{nt} - 1)/\sigma_n\}$ satisfy the CLT condition:

$$\sup\{E_*(W_{n1}^4); n \geq 1\} < \infty,\ E_*(W_{n1}^2 W_{n2}^2) \to 1. \tag{9}$$

Conditions (8) and (9) are satisfied by weights corresponding to the three schemes in the above examples as discussed by Cheng and Huang (2010, p. 2891).

THEOREM 1. *Assume that* (8), (9)*, smoothness conditions SM1-SM4 and moment condition M1 described in the Supplementary material hold. Then for almost all data, as* $n \to \infty$

$$\sigma_n^{-1} n^{1/2}(\hat{\theta}_{*n} - \hat{\theta}_n) \to N\{0, \sigma^2(H)J^{-1}\} \tag{10}$$

in distribution.

Remark 1. Since $0 < 1/\sigma_n < 1/k$, if the sequence $1/\sigma_n$ is bounded below by a positive number, the order of convergence of the bootstrap estimator is the same as that of the original estimator. A lower bound of $1/\sigma_n$ exists for the bootstrap schemes M, E and U in the examples. The standard deviation of the weights σ_n in the denominator of the scaling reflects the contribution of the corresponding weights.

We now develop an algorithm for efficiently computing the estimates for the bootstrap samples. Following on (5), define a function g of two arguments $\tilde{\theta}$ and θ as the sum of conditional expectations

$$g(\tilde{\theta}, \theta) = \sum_{t=1}^{n} E_{\tilde{\theta}} \left([1 - H\{X_t/v_t^{1/2}(\theta)\}]\{\dot{v}_t(\theta)/v_t(\theta)\} \mid \{X_j, j \le t-1\} \right),$$

where the conditional expectation is calculated under the true parameter $\tilde{\theta}$ and for $t = 1$, the σ-field $\{X_j, j \le t-1\}$ denotes the trivial σ-field. When $\tilde{\theta} = \theta$ and is equal to the true parameter θ_{0H},

$$\begin{aligned} g(\tilde{\theta}, \theta) &= \sum_{t=1}^{n} E[1 - H\{X_t/v_t^{1/2}(\theta_{0H})\}]\{\dot{v}_t(\theta_{0H})/v_t(\theta_{0H})\} \\ &= \sum_{t=1}^{n} E[\{1 - H(\epsilon_t/c_H^{1/2})\}\{\dot{v}_t(\theta_{0H})/v_t(\theta_{0H})\}] = 0. \end{aligned}$$

So we assume that a solution to $\sum_{t=1}^{n}[1 - H\{X_t/v_t^{1/2}(\theta)\}]\{\dot{v}_t(\theta)/v_t(\theta)\} = 0$ is close to the solution of

$$\sum_{t=1}^{n}[1 - H\{X_t/v_t^{1/2}(\theta)\}]\{\dot{v}_t(\theta)/v_t(\theta)\} = g(\tilde{\theta}, \theta).$$

Using a Taylor expansion of g as a differentiable function of $\tilde{\theta}$ around θ and replacing $v_t(\theta)$ by $\hat{v}_t(\theta)$, we obtain the following recursive equation for computing the updated estimate $\tilde{\theta}$ of $\hat{\theta}_n$ from the current estimate θ.

$$\tilde{\theta} = \theta - [2/E\{\epsilon \dot{H}(\epsilon)\}] \Big[\sum_{t=1}^{n} \{\dot{\hat{v}}_t(\theta)\dot{\hat{v}}_t(\theta)^{\mathrm{T}}/\hat{v}_t^2(\theta)\} \Big]^{-1} \sum_{t=1}^{n}[1 - H\{X_t/\hat{v}_t^{1/2}(\theta)\}]\{\dot{\hat{v}}_t(\theta)/\hat{v}_t(\theta)\}.$$

Moreover the bootstrap estimate $\hat{\theta}_{*n}$ can be computed by updating $\tilde{\theta}_*$ from the current estimate θ using a similar weighted recursive equation as follows.

$$\tilde{\theta}_* = \theta - [2/E\{\epsilon \dot{H}(\epsilon)\}] \Big[\sum_{t=1}^{n} w_{nt}\{\dot{\hat{v}}_t(\theta)\dot{\hat{v}}_t(\theta)^{\mathrm{T}}/\hat{v}_t^2(\theta)\} \Big]^{-1} \sum_{t=1}^{n} w_{nt}[1 - H\{X_t/\hat{v}_t^{1/2}(\theta)\}]\{\dot{\hat{v}}_t(\theta)/\hat{v}_t(\theta)\}. \tag{11}$$

3. SIMULATION

In our empirical study we aim (i) to assess the validity of (7) and (10) in practice by providing coverage rates of the confidence intervals obtained through the asymptotic normal approximation and bootstrap approximations; and (ii) to provide some empirical evidence for the superior performance of the weighted bootstrap over resampling methods such as percentile t-subsampling and residual bootstrap that are available in the bootstrap literature on GARCH models with heavy-tailed error distributions. In the Supplementary material, we also analyze a recent data on the daily adjusted closing prices of the Nikkei 225 Index of the Japanese market to provide bootstrap estimates of the bias, mean squared error and variance of some M-estimators. Only two M-estimators, namely, the QMLE and least absolute deviation estimator are computed in this limited simulation study. Extensive simulation using other M-estimators will be reported elsewhere.

We generate data from a GARCH $(1,1)$ model with parameters $\omega_0 = 7.62 \times 10^{-6}, \alpha_{01} = 1.54 \times 10^{-1}, \beta_{01} = 8.31 \times 10^{-1}$. We select such parameters since these are the QMLE of the Nikkei data when a GARCH $(1,1)$ model is fitted.

For the error distribution, we choose Student's $t(d)$-distribution with d degrees of freedom, where $d = 3, 4, 5, \infty$. Such choices cover a representative array of errors with $d = 3, 4$ corresponding to heavy-tailed distributions (in the sense of having infinite 4-th moment) and $d = 5, \infty$ corresponding to light-tailed distributions. Similar error distributions were also considered for generating data while studying subsampling schemes by Linton, Pan, and Wang (2010). Since with heavy-tailed distributions the algorithm for computing the QMLE did not always converge for the generated data, coverage rates of the QMLE are reported for light-tailed error distributions while the corresponding rates of the least absolute deviation estimator are reported for all four error distributions.

The sample size for each dataset was $n = 1000$ and we considered $B = 2000$ bootstrap samples. Coverage rates were computed based on $R = 500$ replicates.

We consider four types of bootstrap schemes. The first three types are discussed in Section 2. Type 1: Scheme U with four possible equally-spaced values of a, namely, $a = 0.25, 0.5, 0.75, 1$. Type 2: Scheme M. Type 3: Scheme E which is similar to Scheme U but based on exponential random variables.

Type 4: Percentile t-subsampling: Such bootstrap schemes were considered by Hall and Yao (2003) and Linton, Pan and Wang (2010) for the QMLE. Hidalgo and Zaffaroni (2007) considered a special case which may be called a percentile residual bootstrap. In the sequel, we will call them simply subsampling or residual bootstrap. For this bootstrap, we define residuals by $\tilde{\epsilon}_t = X_t/\hat{v}_t^{1/2}(\hat{\theta}_n)$, $1 \leq t \leq n$ and then the centered residuals by $\hat{\epsilon}_t = \tilde{\epsilon}_t - n^{-1}\sum_{t=1}^{n}\tilde{\epsilon}_t$, $1 \leq t \leq n$. Let the subsample size be m where $m \leq n$. The case $m = n$ is the residual bootstrap. We generate B subsampling bootstrap M-estimates as follows. For each b, $1 \leq b \leq B$, we select a random sample of size m from $\{\hat{\epsilon}_t; 1 \leq t \leq n\}$ and call it $\{\hat{\epsilon}_t^*; 1 \leq t \leq m\}$. We form a new set of data $\{X_t^*, 1 \leq t \leq m\}$ where $X_t^* = \hat{v}_t^{1/2}(\hat{\theta}_n)\hat{\epsilon}_t^*$ and compute $\hat{\theta}_{*n}$ as the b-th subsampling bootstrap M-estimate based on $\{X_t^*; 1 \leq t \leq m\}$. This subsampling or residual bootstrap is not a special case of the weighted bootstrap. We exhibit simulation results with subsampling size m where $m = 0.6n, 0.7n, 0.8n$ and n.

For computing coverage rates under the normal approximation, we use Proposition 3.1 of Mukherjee (2008) to estimate the variance-covariance matrix of the limiting normal distribution needed for the confidence intervals.

The coverage rates of the QMLE and least absolute deviation estimator are reported in Tables 1 and 2. The tables show that for Scheme U, the coverage rates for both QMLE and least absolute deviation estimator are close to each other for different values of a and they are reasonably close to nominal values especially for the intercept parameter ω. Similar comments apply for Scheme M and Scheme E with good coverage rates for both QMLE and least absolute deviation estimator and for all error distributions considered in our study. Moreover, no scheme seems to dominate the other uniformly over various parameters or error distributions in terms of coverage rate.

The coverage rates of the subsampling bootstrap are reasonable for heavy-tailed distributions but they become increasingly poor for both α and β under light-tailed distribution such as $t(5)$ or normal. This is not surprising since such bootstrap schemes were proposed by Hall and Yao (2003) and Linton, Pan and Wang (2010) for heavy-tailed error distributions. The coverage rate of the normal approximation is poor even for the QMLE when the error distribution is normal.

All three weighted bootstrap schemes dominate subampling and residual bootstrap. Our simulation study thus provides some empirical support to prefer the weighted bootstrap approximation over the subsampling schemes or the normal approximation of the distribution of M-estimators when considering GARCH models with a wide spectrum of heavy-tailed as well as light-tailed error distribution.

Compared with some other bootstraps in the literature for GARCH models, the weighted bootstrap is computationally simpler. We can store the core quantities $\{[1 - H\{X_t/\hat{v}_t^{1/2}(\theta)\}]\{\dot{\hat{v}}_t(\theta)/\hat{v}_t(\theta)\}\}$ while computing M-estimates. After that, for each bootstrap replicate, one simply needs to generate weights and solve the equation involving a weighted sum through iteration using (11). Each time, the initial estimate for the iteration is taken to be the M-estimate $\hat{\theta}_n$.

Table 1. The coverage rates of different bootstrap schemes for the QMLE

			ω	α	β	ω	α	β
			95% nominal value			90% nominal value		
$t(5)$ dist.	Scheme U	$a = 0.25$	94.3	91.7	92.2	89.5	87.7	86.2
		$a = 0.5$	96.2	91.5	92.0	91.9	86.8	86.6
		$a = 0.75$	96.6	91.1	92.3	92.7	86.6	87.3
		$a = 1$	96.8	90.8	92.7	93.4	86.1	88.1
	Subsampling	$m = 600$	87.9	92.0	96.3	84.6	86.5	93.6
		$m = 700$	88.6	92.0	96.4	85.7	85.9	93.6
		$m = 800$	88.6	92.7	96.6	85.9	85.8	93.3
		$m = 1000$	89.9	92.1	96.4	86.4	84.2	92.1
	Scheme M		94.6	89.1	90.0	90.9	84.5	85.9
	Scheme E		93.1	88.2	89.3	89.7	83.6	84.7
Normal dist.	Scheme U	$a = 0.25$	90.8	93.3	93.0	80.6	89.9	87.6
		$a = 0.5$	91.7	93.5	93.3	81.9	89.9	88.5
		$a = 0.75$	92.6	93.2	94.0	83.2	89.8	88.9
		$a = 1$	93.5	93.2	94.6	84.3	89.7	89.3
	Subsampling	$m = 600$	92.2	66.6	81.2	87.7	51.0	69.2
		$m = 700$	92.4	59.7	76.2	88.8	45.4	62.6
		$m = 800$	93.2	53.4	69.7	89.9	40.4	56.8
		$m = 1000$	95.6	42.7	60.4	92.4	32.2	44.7
	Scheme M		96.4	93.8	95.1	89.1	90.0	90.5
	Scheme E		95.7	93.1	95.0	87.0	89.7	90.1
	Normal approximation		68.6	92.4	86.4	61.4	87.8	80.2

Table 2. The coverage rates of different bootstrap schemes for the least absolute deviation estimator

			ω	α	β	ω	α	β
			95% nominal value			90% nominal value		
$t(3)$ dist.	Scheme U	$a = 0.25$	95.0	90.4	92.2	90.4	87.3	87.3
		$a = 0.5$	95.0	90.2	91.8	90.7	86.4	87.4
		$a = 0.75$	94.9	90.0	92.0	90.0	86.0	87.2
		$a = 1$	94.8	90.0	91.8	90.1	85.7	87.2
	Subsampling	$m = 600$	86.8	87.5	96.0	82.9	78.2	94.4
		$m = 700$	86.1	87.8	96.4	81.8	77.7	95.2
		$m = 800$	84.5	86.0	97.3	79.8	74.5	95.4
		$m = 1000$	85.3	83.3	97.1	80.7	71.2	95.7
	Scheme M		93.0	87.9	91.8	90.0	85.2	87.5
	Scheme E		93.0	87.5	91.6	90.0	84.6	87.0
$t(4)$ dist.	Scheme U	$a = 0.25$	96.3	90.9	92.4	91.6	86.2	87.7
		$a = 0.5$	96.3	90.6	92.3	92.2	86.0	87.8
		$a = 0.75$	96.7	89.7	92.4	92.7	85.5	87.7
		$a = 1$	97.0	89.3	92.2	93.0	85.1	87.5
	Subsampling	$m = 600$	85.0	80.2	96.5	80.2	68.0	93.7
		$m = 700$	85.7	77.1	96.4	81.7	66.1	93.9
		$m = 800$	87.7	71.5	97.2	83.6	57.2	93.1
		$m = 1000$	86.1	71.0	97.0	80.5	56.2	92.2
	Scheme M		96.6	91.6	93.7	94.5	88.6	89.0
	Scheme E		96.8	91.2	93.5	94.4	88.1	88.6
$t(5)$ dist.	Scheme U	$a = 0.25$	95.9	93.6	93.6	90.7	88.6	88.7
		$a = 0.5$	96.4	93.7	93.8	91.7	88.2	89.7
		$a = 0.75$	96.4	93.4	94.4	92.3	88.2	90.0
		$a = 1$	96.7	93.1	94.7	92.7	87.8	90.3
	Subsampling	$m = 600$	84.5	73.4	95.1	80.5	60.2	90.4
		$m = 700$	85.8	70.7	93.9	81.5	56.1	87.7
		$m = 800$	84.9	64.8	94.1	81.6	50.4	87.6
		$m = 1000$	88.6	56.0	90.3	84.2	42.7	81.6
	Scheme M		96.8	91.7	94.6	93.5	86.9	90.2
	Scheme E		97.0	91.7	94.2	93.6	86.8	89.7
Normal dist.	Scheme U	$a = 0.25$	93.8	93.7	94.9	85.8	89.6	90.0
		$a = 0.5$	94.8	93.5	95.1	87.0	89.1	90.2
		$a = 0.75$	96.1	93.3	95.3	88.2	89.1	90.9
		$a = 1$	96.9	93.3	95.6	89.6	89.3	91.4
	Subsampling	$m = 600$	90.5	52.7	81.4	85.9	40.6	69.0
		$m = 700$	91.7	46.8	75.8	86.4	36.0	62.1
		$m = 800$	91.9	40.9	70.4	87.4	31.6	57.1
		$m = 1000$	93.0	32.6	62.1	88.8	25.4	45.8
	Scheme M		97.7	93.2	95.7	92.1	90.1	91.2
	Scheme E		97.6	92.9	95.4	91.6	90.0	90.9
	Normal approximation		89.0	99.8	88.6	82.0	99.2	84.2

The subsampling bootstrap took a lot longer time than the weighted bootstrap because (a) for each replicate, we need to consider a new set of data and (b) various resulting matrices associated with the new set of data for the computation of the QMLE and least absolute deviation estimator were not invertible. Consequently, those samples were removed and new bootstrap samples were generated.

4. OUTLINE OF PROOF

To derive the asymptotic distribution (10) of the estimator $\hat{\theta}_{*n}$ defined by a root of a smooth function of θ, we use a simple modification of a result of Klimko and Nelson (1978, Theorem 2.1, Corollary 2.1 and Theorem 2.2) as stated in the Supplementary material. Define ρ by $\rho(x) = \int_0^x \psi(t)dt$ for $x \geq 0$ and $\rho(x) = \rho(-x)$ for $x < 0$. Let

$$\hat{m}_t(\theta) = \rho\{X_t/\hat{v}_t^{1/2}(\theta)\} + (1/2)\log \hat{v}_t(\theta),\ \hat{M}_n(\theta) = \sum_{t=1}^{n} \hat{m}_t(\theta),\ \hat{M}_{*n}(\theta) = \sum_{t=1}^{n} w_{nt}\hat{m}_t(\theta).$$

Since $\hat{M}_{*n}(\hat{\theta}_{*n}) = 0$, we verify the conditions of the Klimko and Nelson Theorem for the criterion function $\hat{M}_{*n}$ which is a weighted sum of $\{\hat{m}_t(\theta)\}$. We use following lemmas to show various convergence properties of the weighted sums with respect to the bootstrap distribution.

Let $\{K_t\}$ be a sequence of random variables of the form $K_t = C_{1t} + C_{2t}C_{3t} + C_{4t}C_{5t}C_{6t} + \ldots$ where the sequence of identically distributed random vectors $\{C_t; t \geq 1\}$ is defined by $C_t = (C_{1t}, C_{2t}, \ldots)$ with $C_{it} > 0$ for all $i \geq 1$ and $E(\log^+ C_{i1}) < \infty$.

LEMMA 1. *Let $\{K_t, 1 \leq t \leq n\}$ be independent of $\{w_{nt}\}$ satisfying (8). Then for all $0 < \rho < 1$, $\sigma_n^{-1} \sum_{t=1}^n w_{nt}\rho^t K_t$ converges almost surely.*

The next lemma states the convergence of a bootstrap-weighted average when multiplied by infinitesimal sequence of random variables.

LEMMA 2. *Let $\{a_t\}$ be a stationary ergodic sequence of random variables with $E(|a_1|) < \infty$. Let $\{u_n\}$ be a sequence of random variables (with u_n possibly dependent on $\{a_t, t \leq n\}$) such that $u_n = o_p(1)$ and $\{u_n, a_t, 1 \leq t \leq n\}$ is independent of $\{w_{nt}, 1 \leq t \leq n\}$. Then $(u_n/n)\sum_{t=1}^n w_{nt}a_t = o_p(1)$* almost surely.

LEMMA 3. *Let $\{a_t, 1 \leq t \leq n\}$ be a sequence of second order stationary ergodic random variables independent of $\{w_{nt}; 1 \leq t \leq n\}$ such that $E(a_1) = 0$. Then $(1/n)\sum_{t=1}^n w_{nt}a_t = o_p(1)$ almost surely.*

Let $V = \tilde{a}J$ where $\tilde{a} = E\{(\epsilon/c_H^{1/2})\dot{H}(\epsilon/c_H^{1/2})\}/8 > 0$. We verify conditions of the Klimko and Nelson Theorem by showing that almost surely,

$$n^{-1}\dot{\hat{M}}_{*n}(\hat{\theta}_n) = o_p(1), \tag{12}$$

$$(2n)^{-1}\ddot{\hat{M}}_{*n}(\hat{\theta}_n) \to V, \tag{13}$$

$$\lim_{n\to\infty}\lim_{\delta\to 0+} \sup\{(n\delta)^{-1}|\ddot{\hat{M}}_{*n}(\theta) - \ddot{\hat{M}}_{*n}(\hat{\theta}_n)|; \|\theta - \hat{\theta}_n\| \leq \delta\} < \infty, \tag{14}$$

$$(2\sigma_n n^{1/2})^{-1}\dot{\hat{M}}_{*n}(\hat{\theta}_n) \to N\{0, \sigma^2(H)\, J^{-1}\}. \tag{15}$$

We sketch the common approach for proving (12) - (15) by providing some details of the proofs of (12) and (15) only but leave elaborate arguments for the Supplementary material. For (12), we show that

$$n^{-1}\sup\{|\dot{\hat{M}}_{*n}(\theta) - \dot{M}_{*n}(\theta)|; \theta \in \Theta_0\} = o_p(1), \tag{16}$$

$$n^{-1}\{\dot{M}_{*n}(\hat{\theta}_n) - \dot{M}_{*n}(\theta_{0H})\} = o_p(1), \tag{17}$$

$$n^{-1}\dot{M}_{*n}(\theta_{0H}) = o_p(1). \tag{18}$$

In (16), the difference of two quantities involving $v_t(\theta)$ and $\hat{v}_t(\theta)$ is bounded by quantities of the form $\{\rho^t K_t\}$ and so $|\dot{\hat{M}}_{*n}(\theta) - \dot{M}_{*n}(\theta)|$ is bounded above by the $\{w_{nt}\}$-weighted sum of $\{\rho^t K_t\}$ which by Lemma 1 converges with respect to the bootstrap distribution. In (17), the difference of two quantities are weighted smooth functions of the difference involving $\hat{\theta}_n$ and θ_{0H} and hence goes to zero. Finally, (18) involves weighted sum $\{\sum_{t=1}^{n} w_{nt} a_t\}$ that converges using Lemma 3. For (15), steps analogous to (16) and (17) can be proved similarly and the final step involves showing the convergence in distribution

$$(2\sigma_n n^{1/2})^{-1} \dot{M}_{*n}(\theta_{0H}) \to N\{0, \sigma^2(H)\, J^{-1}\}.$$

For this we verify conditions of Lemma 4.6 of Praestgaard and Wellner (1993) for the asymptotic normality of a weighted sum of the exchangeable random variables $\{w_{nt}/\sigma_n; 1 \le t \le n\}$ since with the function h defined by $h(x) = (1/2)\{1 - H(x)\}$,

$$(2\sigma_n n^{1/2})^{-1} \dot{M}_{*n}(\theta_{0H}) = n^{-1/2} \sum_{t=1}^{n} (w_{nt}/\sigma_n)(1/2) h(\epsilon_t / c_H^{1/2}) \{\dot{v}_t(\theta_{0H}) / v_t(\theta_{0H})\}.$$

ACKNOWLEDGEMENT

I am grateful to Mr Hang Liu for his help with the extensive computation of the paper. I would like to thank the Editor, Associate Editor and the Referees for their careful reading and for making many useful comments to improve the paper. Thanks are also due to Dr Daniel Grose for transforming the computer programs into R-packages.

SUPPLEMENTARY MATERIAL

The Supplementary material contains details of the proof of Theorem 1 elaborating on the 'Outline of proof', proofs of the lemmas, more simulation results explaining the behaviour of the coverage probabilities under the larger sample size $n = 2500$ for the GARCH (1,1) model, simulation results for the GARCH (2,1) model and the computer programs and R-packages for the implementation of the bootstrap methods.

REFERENCES

[1] Berkes, I., Horvath, L. & Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli*. **9**. 201-228.

[2] Chatterjee, S. & Bose, A. (2005). Generalized bootstrap for estimating equations. *Annals of Statistics*. **33**. 414–436.

[3] Cheng, G. & Huang, J. (2010). Bootstrap consistency for general semiparametric M-estimation. *Annals of Statistics*. **38**. 2884-2915.

[4] Hall, P. & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*. **71**. 285-317.

[5] Hidalgo, J. & Zaffaroni, P. (2007). A goodness-of-fit test for ARCH(∞) models. *Journal of Econometrics*. **141**. 973-1013.

[6] Klimko, L. A. & Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics*. **6**. 629-642.

[7] Linton, O., Pan, J. & Wang, H. (2010). Estimation for a nonstationary semi-strong GARCH(1, 1) model with heavy-tailed errors. *Econometric Theory*. **26**. 1-28.

[8] Mak, T. (1993). Solving non-linear estimation equations. *Journal of Royal Statistical Society*. B. **55**. 945-955.

[9] Mak, T., Wong, H. & Li, W. (1997). Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. *Computational Statistics and Data Analysis*. **24**. 169-178.

[10] Mukherjee, K. (2008). M-estimation in GARCH model. *Econometric Theory*. **24**. 1530-1553.

[11] Praestgaard, J. & Wellner, J. (1993). Exchangeably weighted bootstrap of the general empirical process. *Annals of Probability*. **21**. 2053-2086.

[12] Varga, L. & Zempleni, A. (2012). Weighted bootstrap in GARCH models. arXiv: 1209.1302v1.

[*Received on* 2 *January* 2017. *Editorial decision on* 1 *April* 2017]