University of Groningen

Confidence intervals for ARMA-GARCH Value-at-Risk

Spierdijk, Laura

Published in:
Computational Statistics & Data Analysis

DOI:
[10.1016/j.csda.2014.08.011](https://doi.org/10.1016/j.csda.2014.08.011)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Spierdijk, L. (2016). Confidence intervals for ARMA-GARCH Value-at-Risk: The case of heavy tails and skewness. *Computational Statistics & Data Analysis*, 100, 545-559.
<https://doi.org/10.1016/j.csda.2014.08.011>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Confidence intervals for ARMA–GARCH Value-at-Risk: The case of heavy tails and skewness

Laura Spierdijk*

University of Groningen, Faculty of Economics and Business, Department of Economics, Econometrics and Finance, P.O. Box 800, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Article history:

Received 20 December 2013

Received in revised form 14 August 2014

Accepted 16 August 2014

Available online 27 August 2014

Keywords:

Value-at-Risk

ARMA–GARCH

Quasi-maximum likelihood

Subsample bootstrap

ABSTRACT

It is a well-known result that, when the ARMA–GARCH model errors lack a finite fourth moment, the asymptotic distribution of the quasi-maximum likelihood estimator may not be Normal. In such a scenario the conventional bootstrap turns out inconsistent. Surprisingly, simulations show that the conventional bootstrap, despite its inconsistency, provides accurate confidence intervals for ARMA–GARCH Value-at-Risk (VaR) in case of various symmetric error distributions without finite fourth moment. The usual bootstrap does fail, however, in the presence of skewed error distributions without finite fourth moment. In this case several other methods for estimating confidence intervals fail as well. A residual subsample bootstrap is proposed to obtain confidence intervals for ARMA–GARCH VaR. According to theory, this 'omnibus' method produces confidence intervals with asymptotically correct coverage rates under very mild conditions. By means of a simulation study the favorable finite-sample properties of the residual subsample bootstrap are illustrated. Confidence intervals for ARMA–GARCH VaR with good coverage rates are established, even when other methods fail in the presence of skewed model errors without finite fourth moment. The estimation of confidence intervals by means of the residual subsample bootstrap is illustrated in an empirical application to daily stock returns.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

During the past decades Value-at-Risk (VaR) has become a standard tool in risk management. For example, the capital requirements for insurance companies operating in the European Union, as prescribed by the Solvency II Directive, are based on the 99.5% VaR of the portfolio of assets and liabilities measured over a one-year time horizon (Sandström, 2011). Similarly, VaR plays a role in the Basel II and III frameworks for banks.

The ε -VaR is defined as the ε % quantile of a given profit-and-loss (P – L) distribution. VaR estimation requires certain assumptions about the latter distribution. When the P – L distribution of a portfolio of securities is the object of interest, it has become fairly standard to assume a distribution of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) type (Engle, 1982; Bollerslev, 1986). GARCH models capture an important stylized fact of asset returns known as volatility clustering: high (absolute) returns tend to be followed by more high returns and low (absolute) returns tend to be followed by more low returns.

An attractive property of GARCH models is that consistent estimation of the model parameters does not require knowledge of the distribution of the model errors. Quasi-maximum likelihood (QML) is based on the Gaussian likelihood function and yields consistent parameter estimates even when normality is violated, provided that certain regularity

* Tel.: +31 50 363 5929; fax: +31 50 363 7337.

E-mail address: L.Spierdijk@rug.nl.

conditions hold (Hall and Yao, 2003; Francq and Zakoian, 2004). Consistent (but generally inefficient) estimation of the model parameters' covariance matrix is possible using the sandwich estimator proposed by Bollerslev and Wooldridge (1992). QML routines for ARMA–GARCH model estimation are available in software packages such as R, Eviews and Stata.

VaR estimates depend on the estimated parameters of the underlying P – L distribution, which are subject to estimation uncertainty. To judge the accuracy of a VaR estimate, it is crucial to quantify this uncertainty in terms of a confidence interval (Jorion, 1996). That is, an estimated VaR of \$100 million with a 95% confidence interval of \$[95, 105] million indicates much more accuracy than a VaR point estimate of \$100 million with a 95% confidence interval equal to \$[50, 150] million. When the VaR confidence interval is found to be wide, caution is required when decisions are to be based on the inaccurate point estimate. For example, if a VaR estimate is used to set a capital reserve, additional capital on top of the VaR point estimate might be required to account for the underlying estimation uncertainty. Although the point estimates in the aforementioned example are the same, the associated capital reserves should be different due to the differences in estimation uncertainty. Another application of the confidence intervals associated with different VaR point estimates is the selection of the estimate with the lowest percentage of estimation uncertainty.

Several methods have been used to quantify the parameter uncertainty associated with model-based VaR estimates. Among these methods are the Delta method (Duan, 1994), the conventional n -out-of- n with-replacement bootstrap (Dowd, 2002; Christoffersen and Gonçalves, 2005; Pascal et al., 2006; Goa and Song, 2008) and analytical confidence intervals based on the VaR estimator's asymptotic distribution (Goa and Song, 2008; Francq and Zakoian, 2012). These methods have been shown to work well when the fourth moment of the GARCH errors is finite.

Conventional methods to construct confidence intervals generally rely on asymptotic normality and fail when this assumption does not hold. For example, in a general setting the conventional bootstrap is not consistent in the absence of asymptotic normality (Athreya, 1987a,b; Knight, 1989; Hall, 1990; Mammen, 1992). More specifically, Hall and Yao (2003) show that non-normality of the QML estimator and inconsistency of the conventional bootstrap arise in a GARCH model when the fourth moment of the model errors is not finite and the distribution of the squared model errors is not in the domain of attraction of the Normal distribution. Similarly, also for ARMA–GARCH models a finite fourth moment is needed to achieve asymptotic normality (Francq and Zakoian, 2004) and thereby consistency of the conventional bootstrap.

The literature has tried to circumvent the QML estimator's non-normality in several ways. Chan et al. (2007) use QML in combination with the Hill estimator (Hill, 1975) to estimate extreme conditional quantiles, but find that the resulting coverage rates are highly sensitive to the length of the tail sample used by the Hill estimator. Another option is to step away from QML and to use ML based on a specific error distribution (Bams et al., 2005; Pascal et al., 2006; Luger, 2012). As usual, however, imposing additional distributional assumptions incurs the risk of misspecification. Another alternative to QML is the Least Absolute Deviations estimator based on non-smooth optimization (Peng and Yao, 2003; Goa and Song, 2008; Huang et al., 2008). It is also possible to estimate VaR in a fully non-parametric way (Chen and Tang, 2005). However, the latter two approaches are non-standard, which makes them less attractive than QML-based ARMA–GARCH VaR estimation.

Instead of stepping away from the QML estimator, this study proposes a QML-based 'omnibus' approach to estimate confidence intervals for ARMA–GARCH VaR. According to theory, this omnibus method produces confidence intervals with asymptotically correct coverage rates regardless of the asymptotic properties of the QML estimator. Our method is based on the subsample bootstrap of Sherman and Carlstein (2004) and is henceforth referred to as the residual subsample bootstrap. It boils down to an m -out-of- n non-replacement bootstrap applied to the ARMA–GARCH model errors. The conditions imposed on the subsample bootstrap to achieve consistency are much weaker than those required for the conventional n -out-of- n with-replacement bootstrap (Horowitz, 2001). In particular, our approach imposes minimal regularity conditions and allows for heavy-tailed and skewed error distributions, with or without finite fourth moment. The subsample size can be chosen in a data-driven way in the spirit of Hall (1990). In contrast with the subsample bootstrap of Bertail et al. (1999) and Politis and Romano (1994) our subsampling procedure does neither require knowledge nor estimation of the rate of asymptotic convergence. Moreover, the subsample bootstrap is as easy to implement as the ordinary bootstrap and can be conveniently combined with standard software for QML estimation.

By means of a simulation study we first show that the conventional bootstrap, despite its theoretical inconsistency, yields surprisingly accurate confidence intervals in case of symmetric error distributions without finite fourth moment. The conventional bootstrap does fail, however, in the presence of skewed error distributions that lack a finite fourth moment. In this case also the approach of Chan et al. (2007) fails. These findings emphasize that any omnibus method for calculating confidence intervals should have correct coverage rates even when the ARMA–GARCH errors are skewed and lack a finite fourth moment.

We assess the finite-sample properties of the residual subsample bootstrap by means of simulation. For a wide range of symmetric and skewed error distributions with or without finite fourth moment this method produces confidence intervals for ARMA–GARCH VaR with accurate coverage rates, even when other methods fail. We illustrate the estimation of confidence intervals by means of the residual subsample bootstrap in an empirical application involving daily stock returns. To our best knowledge, our study is the first to analyze confidence intervals for QML-based ARMA–GARCH VaR in the presence of skewed model errors without finite fourth moment.

The setup of the remainder of this article is as follows. Section 2 describes how QML-based ARMA–GARCH VaR is estimated. The statistical complications caused by (skewed) model errors without finite fourth moment are discussed in Section 3. Our solution is the residual subsample bootstrap, whose finite-sample properties are analyzed in Section 4 by means of simulation. The estimation of confidence intervals for ARMA–GARCH VaR by means of the residual subsample

bootstrap is illustrated in Section 5 in an application to daily stock returns. Section 6 concludes. An online appendix with supplementary material is available (see Appendix D).

2. Estimation of QML-based ARMA–GARCH VaR

ARMA–GARCH specifications are location-scale models belonging to the class of time-series models with GARCH errors (McNeil and Frey, 2000; Li et al., 2002). Alternative VaR models have been studied by e.g. Engle and Manganelli (2004) and Xiao and Koenker (2006).

For the sake of exposition we confine the analysis to the ARMA(1, 1)–GARCH(1, 1) model, but we emphasize that our approach can be applied to ARMA–GARCH models of any order. We thus consider the following ARMA(1, 1)–GARCH(1, 1) model for $t = 1, \dots, n$:

$$r_t = \mu_t + \sigma_t z_t; \quad (1)$$

$$\mu_t = \mu + \gamma r_{t-1} + \theta(r_{t-1} - \mu_{t-1}); \quad (2)$$

$$\sigma_t^2 = \omega + \alpha(r_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2; \quad (3)$$

$$z_t \sim D(0, 1) \text{ i.i.d.} \quad (4)$$

Here $D(0, 1)$ denotes a distribution with mean 0 and variance 1, which is potentially skewed and/or heavy-tailed. Covariance-stationarity is imposed by assuming that $|\gamma| < 1$, $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$.

Given a time series sample $r_1, \dots, r_n$, we estimate the ARMA–GARCH model by means of QML, yielding the vector of estimated model parameters $\hat{\lambda} = (\hat{\mu}, \hat{\gamma}, \hat{\theta}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$, as well as $\hat{\mu}_t$ and $\hat{\sigma}_t$ for $t = 1, \dots, n$. We will refer to $z_t = (r_t - \mu_t)/\sigma_t$ as the model errors and to $\hat{z}_t = (r_t - \hat{\mu}_t)/\hat{\sigma}_t$ as the model residuals ($t = 1, \dots, n$).

With extreme losses as the object of interest, the focus is usually on the left tail of the P – L distribution. Because the distribution's right tail reflects losses in case of short-selling, we consider extreme quantiles in both the left and right tail of the P – L distribution. The $\varepsilon\%$ one-step ahead conditional VaR, denoted $\xi_{n|n+1}^\varepsilon$, satisfies $\mathbb{P}(r_{n+1} < \xi_{n|n+1}^\varepsilon | \mathcal{I}_n) = \varepsilon/100$, where $\mathcal{I}_n$ denotes all information available up to time n . Conditional on $\mathcal{I}_n$, the $\varepsilon\%$ quantile of the return distribution equals

$$\xi_{n|n+1}^\varepsilon = \mu_{n+1} + \sigma_{n+1} Q_z^\varepsilon, \quad (5)$$

where Q_z^ε denotes the $\varepsilon\%$ quantile of the model errors z_t . A point estimate $\hat{\xi}_{n|n+1}^\varepsilon$ of $\xi_{n|n+1}^\varepsilon$ is obtained by replacing μ_{n+1} and σ_{n+1} in Eq. (5) by their forecasted counterparts ($\hat{\mu}_{n+1}$ and $\hat{\sigma}_{n+1}$) and Q_z^ε by an estimate ($\hat{Q}_z^\varepsilon$). Here we consider static forecasts $\hat{\mu}_{n+1}$ and $\hat{\sigma}_{n+1}$, which are obtained by iterating Eqs. (2) and (3) one step ahead using $\hat{\lambda}$, r_n , $\hat{\mu}_n$ and $\hat{\sigma}_n$. We define $\hat{Q}_z^\varepsilon$ as the $\varepsilon\%$ sample quantile of $\hat{z}_1, \dots, \hat{z}_n$ (Christoffersen and Gonçalves, 2005; Goa and Song, 2008). The resulting QML-based VaR estimate is subject to three sources of estimation uncertainty. First, there is parameter uncertainty via $\hat{\mu}_{n+1}$, $\hat{\sigma}_{n+1}$ and $\hat{Q}_z^\varepsilon$. Second, the VaR estimate is subject to sampling uncertainty in $\hat{Q}_z^\varepsilon$. Finally, the estimation uncertainty in $\hat{Q}_z^\varepsilon$ is correlated with that in $\hat{\mu}_{n+1}$ and $\hat{\sigma}_{n+1}$ (Goa and Song, 2008). Because of the various sources of estimation uncertainty it is important to provide a confidence interval in addition to the VaR point estimate (Jorion, 1996).

A well-known limitation of VaR as a risk measure is that it does not satisfy the sub-additivity property. Although other risk measures that do satisfy the sub-additivity property may be preferable from a theoretical point of view, VaR remains among the most widely used risk measures. We therefore limit our analysis to VaR, but we emphasize that it is straightforward to extend the approach to alternative risk measures.

3. QML-based confidence intervals for ARMA–GARCH VaR

Consider the situation that we want to construct a QML-based confidence interval for ARMA–GARCH VaR, knowing that (1) the possibly skewed ARMA–GARCH model errors lack a finite fourth moment and (2) the distribution of the squared model errors is not in the domain of attraction of the Normal distribution. As proven by Hall and Yao (2003), the QML estimator of the model parameters has a non-Normal asymptotic distribution in such a scenario. Consequently, analytical confidence intervals based on asymptotic normality are not applicable (see e.g. Goa and Song, 2008). Moreover, as shown by Hall and Yao (2003), the conventional n -out-of- n with-replacement bootstrap is generally not consistent in such a case. The conventional bootstrap is inconsistent because it does not accurately model the relation among extreme order statistics in the sample. For example, the probability that the $k > 2$ largest values in a bootstrap sample are equal does not converge to 0, whereas this probability converges to 0 for the observed sample itself (Hall and Yao, 2003, p. 297). Yet in the case of heavy-tailed error distributions the QML estimator's limiting distribution is determined by the behavior of extreme order statistics (Hall and Yao, 2003).

3.1. The conventional bootstrap

As already observed by Kinatader (1992), inconsistency of the conventional bootstrap does not necessarily imply that it does not work in practice. Surprisingly, the VaR literature has not yet analyzed the performance of the conventional bootstrap

Table 1
Simulated coverage rates.

<i>n</i>	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
<i>Conventional bootstrap</i>								
500	0.87	0.90	0.84	0.86	0.86	0.68	0.78	0.22
1000	0.90	0.91	0.89	0.88	0.87	0.66	0.72	0.24
5000	0.90	0.91	0.89	0.90	0.89	0.72	0.73	0.27
<i>Normal approximation method</i>								
500	0.91	0.78	0.82	0.83	0.84	0.44	0.55	0.06
	40	40	70	80	50	40	30	110
1000	0.90	0.79	0.83	0.83	0.84	0.40	0.49	0.10
	70	70	130	120	90	60	70	60
5000	0.87	0.78	0.78	0.78	0.74	0.51	0.40	0.09
	300	280	560	420	300	20	260	20

Notes: For different sample sizes (n) and distributions of the GARCH(1, 1) errors z_t (see below), the first part of this table displays the simulated coverage rates of the 90% confidence intervals based on the conventional bootstrap (with $B = 1000$ bootstrap replications). The second part displays the most favorable coverage rates realized by the Normal approximation method. The tail sample sizes for which the coverage rates are achieved are shown in italics. The tail sample sizes that give the most favorable coverage rates have been selected using a grid search over values between 10 and 250 (with step size 10) for $n = 500$, 1000 and between 20 and 600 (step size 20) for $n = 5000$. The VaR quantile is based on $\varepsilon = 1\%$ (i.e., 1% VaR). The underlying pure GARCH(1, 1) model has parameters $\lambda = (0, 0, 0, 0.95 \times 20^2/252, 0.15, 0.8)$.

Case 1: $z_t \sim N(0, 1)$; Case 2: z_t skewed Normal, with skewness parameter $\xi = 5$ (positive skewness); Case 3: z_t^2 in the domain of attraction of a stable distribution with parameter $\alpha = 1.5$; Case 4: z_t with a symmetric Pareto distribution; Case 5: z_t with a Student's t distribution with 3 degrees of freedom; Case 6: z_t with a skewed Student's t distribution with 3 degrees of freedom and skewness parameter $\xi = 5$ (positive skewness); Case 7: z_t equal to a scaled version of $\exp(0.5L_t)$, where L_t is a Laplace distribution with parameter $b = 2/3$ (positive skewness); Case 8: z_t with a Pareto-type distribution (positive skewness). All distributions have been scaled such that z_t has mean zero and unit variance. The distributions of z_t in Cases 1 and 2 have a finite fourth moment. In the other cases z_t lacks a finite fourth moment. Cases 1, 3, 4, and 5 correspond with symmetric error distributions. Cases 2, 6, 7, and 8 apply to skewed distributions.

when it is known to be inconsistent. This section investigates this issue by means of simulation. We focus on the n -out-of- n with-replacement residual bootstrap of Christoffersen and Gonçalves (2005) and Goa and Song (2008), to which we will refer as the 'conventional bootstrap'. A detailed algorithmic description of the conventional bootstrap can be found in Appendix A.

We consider a pure GARCH(1, 1) model with $\lambda = (0, 0, 0, 0.95 \times 20^2/252, 0.15, 0.8)$. This GARCH model features high volatility persistence, with an expected annual volatility of 20%. We focus on the 1% VaR, i.e., we choose $\varepsilon = 1\%$. We consider eight different error term distributions, of which two possess a finite fourth moment (Cases 1 and 2). Among the six distributions without finite fourth moment, three are symmetric (Cases 3, 4, and 5) and three are skewed (Cases 6, 7, and 8). In Cases 3–8 the distribution of the squared model errors is not in the domain of attraction of the Normal distribution. We thus know from Hall and Yao (2003) that the conventional bootstrap is inconsistent. Several distributions that we consider have been shown to be empirically relevant for modeling asset returns, in particular the (skewed) t and Pareto distributions (McNeil and Frey, 2000; Bams et al., 2005; Grabchak and Samorodnitsky, 2010). The details of the eight distributions considered are given in the appendix with supplementary material (see Appendix D).

For a given distribution of z_t , we randomly draw values $z_1, \dots, z_n$ from the relevant distribution and recursively simulate $N = 1000$ time series $r_1, \dots, r_n$ on the basis of Eqs. (1)–(3) and a chosen parameter vector λ . We estimate $\xi_{n|n+1}^\varepsilon$ as outlined in Section 2 and obtain the associated 90% confidence intervals using the conventional bootstrap (see Appendix A). We calculate the true value of $\xi_{n|n+1}^\varepsilon$ using the true values of Q_z^ε and the true static values of μ_{n+1} and σ_{n+1} , derived from the simulated returns $r_1, \dots, r_n$ and the true parameter vector λ . The simulated coverage rate is calculated as the fraction of simulations for which the true value of $\xi_{n|n+1}^\varepsilon$ falls within the estimated confidence interval. The appendix with supplementary material provides further computational details (see Appendix D).

The upper part of Table 1 provides the simulated coverage rates for the conventional bootstrap, based on the sample sizes $n = 500, 1000, 5000$ and the aforementioned eight error distributions. When the model errors possess a finite fourth moment, the conventional bootstrap results in accurate coverage rates for both symmetric and skewed distributions (Cases 1 and 2 in Table 1). Surprisingly, the confidence intervals based on the conventional bootstrap also have accurate coverage rates for various symmetric error distributions without finite fourth moment (Cases 3, 4, and 5 in Table 1). Hence, the conventional bootstrap seems to work properly in the latter cases, despite its theoretical inconsistency. Totally different results are found for the skewed error distributions without finite fourth moment (Cases 6, 7 and 8 in Table 1), for which the conventional bootstrap produces coverage rates that are far from nominal. The simulation results are robust against changes in the (ARMA-)GARCH specification and the other parameters.

The coverage rate is informative about the *unconditional* performance of the bootstrap, because it is an average over several simulated samples. We have also inspected the *conditional* performance of the conventional bootstrap; that is, for specific simulated samples. Also the conditional performance of the bootstrap does not reveal anything wrong, according to an analysis in the spirit of Sen et al. (2010).

Although the good performance of the conventional bootstrap in the case of symmetric error distributions without finite fourth moment might look surprising, similar results have been found by Kinatader (1992). Also Kinatader (1992) considers a situation in which the conventional bootstrap is known to be inconsistent. This is the situation of bootstrapping confidence

intervals for the mean of a random variable whose variance does not exist (see [Athreya, 1987a,b](#)). Although this context is unrelated to ARMA–GARCH models or VaR, it will nevertheless prove helpful in understanding our own findings. The bootstrap is *inconsistent* when used to obtain confidence intervals for the mean of random variables without finite variance ([Athreya, 1987a,b](#)). This implies that the bootstrap statistic's distribution does not converge to the original statistic's limiting distribution. Surprisingly, [Kinateder \(1992\)](#) finds that the empirical coverage rates based on the inconsistent bootstrap are very accurate. Hence, despite its theoretical inconsistency, the bootstrap limiting distribution is apparently not far-off the true limiting distribution. [Kinateder \(1992\)](#) concludes that, although the bootstrap is inconsistent, it does *work* in practice. However, the inconsistent bootstrap is only shown to work well for *symmetric* distributions without finite variance (where the analysis is limited to distributions in the domain of attraction of an α -stable distribution, with $\alpha < 2$). It seems that we have a similar situation. We consider several symmetric distributions without finite fourth moment, while we know from [Hall and Yao \(2003\)](#) that the conventional bootstrap is inconsistent. Yet we find that the theoretically invalid bootstrap actually works in practice, in the sense that it produces confidence intervals with accurate coverage rates. Like in the case of [Kinateder \(1992\)](#) this suggests that the practical consequences of the traditional bootstrap's inconsistency are limited in the symmetric case.

In statistics simulations are generally used to investigate the finite-sample properties of methods that are known to be asymptotically valid. Here we have a different situation: the simulations tell us that in some cases the conventional bootstrap has good finite-sample properties despite its inconsistency. Of course, we cannot use the simulations results to derive a general rule about when it is 'safe' or not to apply the inconsistent conventional bootstrap for estimating confidence intervals for ARMA–GARCH VaR. Nevertheless, the simulation results emphasize how important it is to evaluate the accuracy of confidence intervals for ARMA–GARCH VaR with respect to skewed error distributions that lack a finite fourth moment. The existing literature on QML-based confidence intervals for ARMA–GARCH VaR has, to our best knowledge, only considered symmetric error distributions.

3.2. Normal approximation and data tilting

[Chan et al. \(2007\)](#) provide an alternative QML-based method for obtaining confidence intervals for GARCH VaR. Importantly, their approach does not require the GARCH error distribution to possess a finite fourth moment. Their VaR estimator is slightly different from the one proposed in Section 2 and obtained using the two-step procedure of [McNeil and Frey \(2000\)](#). First, the parameters of the pure GARCH model are estimated by means of QML, after which the Hill estimator ([Hill, 1975](#)) is used to estimate the tail index of the standardized return distribution. The $\varepsilon\%$ VaR estimate is obtained using Eq. (5), where the estimate of Q_z^ε is based on the Hill estimator applied to the model errors. By construction, this approach is only valid for values of ε close to 0 or 1. [Chan et al. \(2007\)](#) use the asymptotic normality of the Hill estimator to construct analytical confidence intervals for their VaR estimator. We will refer to this approach in the sequel as the 'Normal approximation' method. One of the crucial assumptions underlying this method is that the Hill estimator is asymptotically unbiased. More details about the Normal approximation approach and its assumptions are given in the appendix with supplementary material (see [Appendix D](#)).

[Chan et al. \(2007\)](#) show that the coverage rates of the VaR confidence intervals are highly sensitive to the choice of the 'tail sample size' used for the calculation of the Hill estimator. Yet for finite samples it is not clear how to choose it optimally; see for example [Huisman et al. \(2001\)](#) for a discussion.

We run the same simulation study as in Section 3.1, but now applied to the Normal approximation method. We calculate coverage rates for a range of tail sample sizes. The lower part of [Table 1](#) displays the most favorable coverage rates for the confidence intervals based on the Normal approximation method, together with the tail sample size for which the best coverage rates are achieved. The coverage rates have poor accuracy in case of skewed error distributions without finite fourth moment. Even for some symmetric distributions the coverage probability error turns out substantial. In particular, the drop in coverage rates when the sample size increases to $n = 5000$ deserves some attention. This phenomenon is likely to be caused by the bias in the Hill estimator, which is ignored in the construction of the confidence interval. It is not observed for all parameter values though. The results in [Table 1](#) illustrate that, for fixed sample size n , the optimal tail sample size varies substantially across distributions. The highest coverage rates found are only feasible when the optimal tail sample size is known upfront, which is not the case in practice. This finding confirms the difficulty of selecting the optimal tail sample size in practice, when the error distribution is unknown.

Besides the Normal approximation method, [Chan et al. \(2007\)](#) also propose data tilting to obtain confidence intervals for their VaR estimator ([Hall and Yao, 2003](#); [Peng and Qi, 2006](#)). Like the Normal approximation method, this method is sensitive to the choice of the tail sample size. [Chan et al. \(2007\)](#) show that the confidence intervals based on the Normal approximation are superior in terms of coverage rates. They explain this from the sensitivity of the data tilting method to the use of model residuals instead of true model errors. Our simulations confirm the superior performance of the Normal approximation method in comparison with the data tilting approach. We therefore only report estimation results for the Normal approximation method. The simulation results for the data tilting method are available upon request.

4. Residual subsample bootstrap

The previous section has made clear that the conventional bootstrap and the methods of [Chan et al. \(2007\)](#) can fail in the presence of skewed model errors without finite fourth moment. The goal of this section is to illustrate the omnibus nature of

the residual subsample bootstrap. By means of a simulation study we investigate its finite-sample properties. We show that the residual subsample bootstrap produces confidence intervals with good coverage rates for a wide range of ARMA–GARCH specifications, even when other methods fail in the presence of skewed model errors without finite fourth moment. We start with a description of the residual subsample bootstrap, followed by some simulation results. The simulation-based performance evaluation of our subsampling method will explicitly consider skewed error distributions without finite fourth moment.

4.1. Methodology

Given a stationary and possibly dependent sample $r_1, \dots, r_n$, Sherman and Carlstein (2004) consider a general statistic $s_n = s_n(r_1, \dots, r_n)$ as a consistent point estimator of a fixed target parameter s . Their subsample bootstrap is based on subsampling without replacement. It assumes that $a_n(s_n - s)$ converges in distribution to a non-degenerate limiting distribution for $n \rightarrow \infty$ and some sequence $a_n \rightarrow \infty$, but does neither require knowledge of (a_n) nor of the limiting distribution. The latter property is a unique feature of the subsample bootstrap of Sherman and Carlstein (2004), which is not shared by other subsample bootstrap methods (cf. Bertail et al., 1999; Politis and Romano, 1994). We notice that the subsample bootstrap proposed by Bertail et al. (1999) and Politis and Romano (1994) is valid under the same conditions as the subsample bootstrap proposed by Sherman and Carlstein (2004). However, the two bootstrap procedures are different. The latter subsample bootstrap does neither require knowledge of (a_n) nor of the limiting distribution, while the former needs known or estimated values of (a_n) .

Apart from the requirement regarding the limiting distribution, two other minimal conditions for the subsample bootstrap's consistency are imposed. We assume that (1) the subsample size $\ell = \ell(n)$ satisfies $\ell(n) \rightarrow \infty$, $\ell(n)/n \rightarrow 0$, and $a_\ell/a_n \rightarrow 0$ for $n \rightarrow \infty$ and (2) the mixing condition as described in Sherman and Carlstein (2004, Condition R.3, p. 127) is satisfied. These conditions are much weaker than those required for consistency of the conventional n -out-of- n with-replacement bootstrap (Horowitz, 2001). Sherman and Carlstein (2004) prove that the subsample bootstrap yields confidence intervals with asymptotically correct coverage rates. They also show that the width of the confidence interval converges in probability to zero for $n \rightarrow \infty$.

Consider the situation that the ARMA–GARCH model errors lack a finite fourth moment and that the distribution of the squared model errors is not in the domain of attraction of the Normal distribution. We already mentioned the result that the conventional bootstrap's inconsistency in this scenario is due to the bootstrap's inability to accurately model the relation among extreme order statistics in the sample (Hall and Yao, 2003). Yet in the case of heavy-tailed error distributions the QML estimator's limiting distribution is determined by the behavior of extreme order statistics, causing the conventional bootstrap to be inconsistent.

In general, non-replacement subsampling methods – such as our residual subsample bootstrap based on Sherman and Carlstein (2004) – are consistent even when the conventional bootstrap fails in the heavy-tailed case. Subsampling works because the resulting subsamples are random samples of the true population distribution, instead of estimators of the latter distribution (Horowitz, 2001). Consequently, subsampling methods succeed in accurately modeling the relation among extreme order statistics in the sample, in contrast to the conventional bootstrap. For example, the probability that the $k > 2$ largest values in a bootstrap subsample are equal converges to 0 for the m -out-of- n bootstrap, provided that $m/n \rightarrow 0$ (Hall and Yao, 2003, p. 297). The formal proof that the subsample bootstrap of Sherman and Carlstein (2004) yields asymptotically correct coverage probabilities relies heavily on the related work of Politis and Romano (1994) and Bertail et al. (1999) and is not repeated here.

The price that is paid for the omnibus nature of the subsample bootstrap is inefficiency: the resulting confidence intervals typically have a length of order a_ℓ^{-1} instead of a_n^{-1} (where ℓ denotes the subsample size and n the full sample size). Yet the main problem with the QML estimator of ARMA–GARCH models whose errors lack a finite fourth moment is that the range of possible limiting distributions and associated convergence rates is extraordinarily large (Hall and Yao, 2003). Consequently, the VaR estimator's limiting distribution and speed of convergence are typically unknown in practice, due to which fully efficient confidence intervals are infeasible. It is therefore attractive to use the subsample bootstrap for confidence interval estimation despite its inefficiency.

We propose a *residual* subsample bootstrap for estimating QML-based confidence intervals for ARMA–GARCH VaR. This bootstrap is an application of the subsample bootstrap of Sherman and Carlstein (2004) to the ARMA–GARCH model errors z_t . As usual the model errors are assumed to be independent and identically distributed (i.i.d.), in which case the mixing condition is satisfied. In this setting the object of interest is the conditional quantile $\xi_{n|n+1}^\varepsilon$, which boils down to a function of λ and Q_z^ε (the fixed parameters), as well as $r_1, \dots, r_n$ (the observed sample). We can thus write $\xi_{n|n+1}^\varepsilon = R(r_1, \dots, r_n; \lambda, Q_z^\varepsilon)$, for some function $R(\cdot)$. To achieve consistency, the subsample bootstrap requires much weaker assumptions than the conventional bootstrap. We assume that $a_n(\hat{\xi}_{n|n+1}^\varepsilon - \xi_{n|n+1}^\varepsilon)$ converges in distribution to a non-degenerate limiting distribution for $n \rightarrow \infty$ and some unknown sequence $a_n \rightarrow \infty$. We also assume that the aforementioned condition (1) holds. Because the model errors are assumed to be i.i.d., they automatically satisfy the mixing condition (2). We refer to Appendix B for a technical remark about the applicability of the bootstrap of Sherman and Carlstein (2004) – designed for a fixed target parameter – to conditional quantiles.

The resulting residual subsample procedure works as follows. We first fit an ARMA–GARCH model to the observed sample $r_1, \dots, r_n$ by means of QML, resulting in the vector of estimated ARMA–GARCH parameters $\hat{\lambda}$, $\hat{Q}_z^\varepsilon$, as well as $\hat{\mu}_t$ and $\hat{\sigma}_t$ for

$t = 1, \dots, n$. A random subsample of length ℓ is drawn *without* replacement from the ARMA–GARCH model residuals $\tilde{z}_1, \dots, \tilde{z}_n$, yielding $\tilde{z}_1, \dots, \tilde{z}_\ell$. Next, $\tilde{\lambda}$ and $\tilde{z}_1, \dots, \tilde{z}_\ell$ are used to recursively generate values $\tilde{r}_1, \dots, \tilde{r}_\ell$ according to Eqs. (1)–(3). The bootstrap sample $\tilde{r}_1, \dots, \tilde{r}_\ell$ is used to estimate the ARMA–GARCH model by means of QML, yielding $\tilde{\lambda}$, as well as $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ for $t = 1, \dots, \ell$. $\tilde{Q}_\varepsilon$ is defined as the $\varepsilon\%$ sample quantile of $\tilde{z}_t = (\tilde{r}_t - \tilde{\mu}_t)/\tilde{\sigma}_t$, $t = 1, \dots, \ell$ and used as a bootstrap estimator of Q_ε . $\tilde{\lambda}$ and the observed $r_1, \dots, r_n$ are used to statically calculate $\tilde{\mu}_{t+1}^*$ and $\tilde{\sigma}_{t+1}^*$ for $t = 1, \dots, n$, according to Eqs. (2)–(3). Finally, $\tilde{\xi}_{n|n+1}^\varepsilon$ is calculated from Eq. (5) using $\tilde{\mu}_{n+1}^*$, $\tilde{\sigma}_{n+1}^*$, and $\tilde{Q}_\varepsilon$. These steps are repeated B times, resulting in $\tilde{\xi}_{n|n+1,1}^\varepsilon, \dots, \tilde{\xi}_{n|n+1,B}^\varepsilon$. The proposed $x\%$ confidence interval has the form $\left[\tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1+x)/2}, \tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1-x)/2} \right]$. Here Q_ξ^y is defined as the empirical $y\%$ quantile of $\tilde{\xi}_{n|n+1,1}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon, \dots, \tilde{\xi}_{n|n+1,B}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon$, with $\hat{\xi}_{n|n+1}^\varepsilon$ calculated from $\hat{\mu}_{n+1}$, $\hat{\sigma}_{n+1}$, and $\hat{Q}_\varepsilon$ according to Eq. (5). A more detailed step-by-step algorithmic description of the residual subsample bootstrap can be found in Appendix C.

Three main differences emerge from a comparison of the conventional bootstrap (Appendix A) and the residual subsample bootstrap (Appendix C): (1) The subsample bootstrap is based on sampling with replacement; (2) It resamples $\ell < n$ instead of n observations; (3) The confidence intervals are constructed using a slightly different percentile method than the one commonly used for the conventional bootstrap (see e.g. Efron and Tibshirani, 1993).

How to choose the subsample size in practice? Intuitively, a larger subsample size results in bootstrap estimators that are based on more data. Estimators based on more data are subject to less parameter uncertainty, resulting in a smaller expected confidence interval width. On the other hand, with a larger subsample size the subsamples tend to overlap more. At the same time, $\tilde{\xi}_{n|n+1,1}^\varepsilon$ and $\tilde{\xi}_{n|n+1}^\varepsilon$ become more dependent. These two effects result in a higher coverage probability error (CPE). Hence, there is a trade-off between expected confidence interval width and CPE. On the basis of a simulation study, Sherman and Carlstein (2004) recommend a subsample size between (a constant times) $n^{1/2}$ and (a constant times) $n^{2/3}$. Our simulations and empirical analysis will therefore consider subsample sizes equal to $\ell_1 = c_1 n^{1/2}$, $\ell_2 = c_2 n^{1/2}$, $\ell_3 = (c_2 n^{1/2} + c_3 n^{2/3})/2$ and $\ell_4 = c_3 n^{2/3}$ for $c_1 = 2.5$, $c_2 = 3.5$, and $c_3 = 2.0$. To strike a proper balance between coverage accuracy and interval width, we ideally choose the largest value of ℓ such that the CPE falls below a given threshold $\delta > 0$. For empirical applications Sherman and Carlstein (2004) recommend a data-driven choice of the subsample size by selecting the largest subsample size (over a grid of subsample sizes) satisfying $|\mathbb{P}^*(\tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1+x)/2} \leq \hat{\xi}_{n|n+1}^\varepsilon \leq \tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1-x)/2}) - x| < \delta$, where x is the confidence level and $\mathbb{P}^*(\cdot)$ denotes the probability measure under bootstrap resampling. This method of choosing the subsample size is based on a bootstrap-based estimate of the coverage probability, using observed data only (see e.g. Example 1.3 and Section 3.7 in Hall, 1992). The computational details of the residual subsample bootstrap can be found in the appendix with supplementary material (see Appendix D).

Our residual subsample bootstrap applies the subsample bootstrap of Sherman and Carlstein (2004) to the ARMA–GARCH model residuals. Alternatively, we could apply the subsample bootstrap to overlapping blocks of raw data; see Sherman and Carlstein (2004). Our residual subsample bootstrap is a parametric or model-based subsample bootstrap, while the subsample block bootstrap is known as a non-parametric type of subsampling. Notice that we cannot apply the conventional block-based bootstrap methods to ARMA–GARCH models whose model errors do not have a finite fourth moment and whose squared model errors are not in the domain of attraction of the Normal distribution. That is, as explained by Paparoditis and Politis (2002, p. 143), the conventional block-based bootstrap methods are limited to linear or approximately linear statistics that are root- n consistent and asymptotically Normal. Already the requirements of root- n consistency and asymptotic normality are not satisfied in this setting.

In an early modeling stage we have done some simulations using the subsample block bootstrap, which led to the following conclusions. The subsample block bootstrap turns out relatively inaccurate in terms of coverage rates. Pre-whitening the raw data by imposing a parametric ARMA–GARCH model leads to more efficient estimation and improves the numerical accuracy of the subsample bootstrap. This holds particularly true when the raw data contain a lot of autocorrelation. In this case the block length has to be chosen relatively large. Consequently, the number of ‘independent’ non-overlapping blocks (essentially n/ℓ) is relatively low, due to which the block bootstrap turns out relatively inaccurate. The performance difference between parametric and non-parametric bootstrap methods is a well-known phenomenon in the bootstrap literature. For example, Horowitz (2001, p. 3185) points out that, in general, the numerical accuracy of the parametric bootstrap tends to be much higher than that of the non-parametric bootstrap. Our initial simulation results indicate that the same holds for parametric and non-parametric subsampling.

4.2. Simulation results

We consider various ARMA–GARCH models, for sample sizes $n = 500, 1000, 5000$. In all specifications the distribution of the model errors is chosen from the eight distributions used in the simulations of Section 3. We report detailed simulations results for three specific ARMA–GARCH models. Model 1 is the pure GARCH(1, 1) model considered in Section 3. It has parameter vector $\lambda = (0, 0, 0, 0.95 \times 20^2/252, 0.15, 0.8)$, corresponding with relatively high volatility persistence and an average expected annual volatility of 20%, while the skewed distributions under consideration feature positive skewness. We take $\varepsilon = 1\%$. Model 2 is an ARMA(1, 1)–GARCH(1, 1) model with parameter vector $\lambda = (0, 0.5, 0.3, 0.95 \times 20^2/252, 0.4, 0.55)$, corresponding with relatively low volatility persistence. The skewed distributions exhibit positive skewness. For this model we choose $\varepsilon = 5\%$. Model 3 has parameter vector $\lambda = (0, 0.5, 0.3, 0.95 \times 20^2/252, 0.15, 0.8)$.

Table 2

Simulated coverage rates: residual subsample bootstrap (Model 1).

n	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
$\ell = \ell_1$								
500	0.88	0.88	0.88	0.88	0.88	0.85	0.87	0.75
1000	0.90	0.89	0.87	0.85	0.90	0.88	0.90	0.85
5000	0.88	0.89	0.87	0.87	0.89	0.87	0.89	0.91
$\ell = \ell_2$								
500	0.87	0.87	0.84	0.85	0.88	0.86	0.86	0.75
1000	0.86	0.88	0.82	0.82	0.86	0.87	0.89	0.84
5000	0.87	0.90	0.88	0.85	0.88	0.88	0.87	0.88
$\ell = \ell_3$								
500	0.85	0.86	0.80	0.84	0.84	0.84	0.87	0.74
1000	0.85	0.85	0.82	0.82	0.84	0.85	0.86	0.78
5000	0.86	0.88	0.87	0.88	0.85	0.86	0.87	0.88
$\ell = \ell_4$								
500	0.83	0.85	0.77	0.76	0.79	0.82	0.83	0.71
1000	0.84	0.83	0.80	0.83	0.83	0.82	0.89	0.77
5000	0.84	0.88	0.86	0.85	0.85	0.88	0.88	0.88

Notes: For different sample sizes (n), subsample sizes (ℓ) and distributions of the GARCH(1, 1) errors z_t , this table displays the simulated coverage rates for 90% confidence intervals based on the residual subsample bootstrap (based on $B = 1000$ bootstrap replications). The VaR quantile is based on $\varepsilon = 1\%$ (i.e., 1% VaR). The underlying GARCH(1, 1) model has parameters $\lambda = (0, 0.5, 0.3, 0.95 \times 20^2/252, 0.15, 0.8)$. The distributions for z_t corresponding with Cases 1–8 are the same as in Table 1. For $n = 500$ the subsample sizes equal $\ell_1 = 56$, $\ell_2 = 78$, $\ell_3 = 102$ and $\ell_4 = 156$, respectively. For $n = 1000$ they are 79, 111, 155 and 200. Finally, for $n = 5000$ they are equal to 177, 247, 416 and 585.

This model features relatively high volatility persistence. The skewed distributions feature negative skewness. We use this model to estimate the 99% VaR; i.e., we choose $\varepsilon = 99\%$.

For sample sizes $n = 500, 1000$ we aim for a maximum CPE of $\delta = 10\%$, while for $n = 5000$ we would like to realize a CPE below 5%. The simulated coverage rates for Models 1, 2 and 3 are displayed in Tables 2–4, for subsample sizes ℓ_1, ℓ_2, ℓ_3 and ℓ_4 . The simulation results show that for the smaller subsample sizes the CPE is almost always below the chosen threshold value, while it occasionally exceeds this value slightly for the larger subsample sizes. In particular, the residual subsample bootstrap achieves accurate coverage rates in case of the skewed distributions without finite fourth moment (Cases 6, 7 and 8). The skewed distribution considered in Case 8 turned out extremely problematic for the conventional bootstrap, as well as for the Normal approximation and data tilting methods.

For the residual subsample bootstrap it is much less of a problem. The chosen CPEs are occasionally exceeded, but only for the smallest sample sizes or the largest subsample sizes. All in all, the simulation results illustrate the omnibus nature of the residual subsample bootstrap. More simulation results (confirming the aforementioned conclusions) can be found in Table 1 (low expected volatility) and Table 2 (high expected volatility) of the appendix with supplementary material (see Appendix D).

The coverage rates are fairly robust to the choice of the subsample size. As mentioned in Section 4, however, smaller subsample sizes result in wider expected confidence intervals. For example, for Model 1 in Case 6 (where the model errors come from a skewed Student's t distribution) with $n = 500$ the average interval widths over the simulation runs are 93% (ℓ_2), 82% (ℓ_3) and 71% (ℓ_4) of the average interval length based on the smallest subsample size ℓ_1 . Given the maximum allowed CPEs of $\delta = 0.10$ ($n = 500, 1000$) and $\delta = 0.05$ ($n = 5000$), our simulations suggest that ℓ_3 generally strikes a good balance between confidence interval width and CPE.

5. Empirical application

Our simulations have shown that the residual subsample bootstrap is able to produce confidence intervals with accurate coverage rates even when the skewed model errors lack a finite fourth moment, while other methods fail in such cases. This section contains an empirical application to daily stock returns to illustrate the use of the residual subsample bootstrap for estimating confidence intervals. We start with a description of a sample of stock returns and the ARMA–GARCH model that we fitted to these returns. Subsequently, we estimate QML-based ARMA–GARCH VaR over a rolling window. Next, we apply the residual subsample bootstrap to estimate confidence intervals, using a data-driven selection of the subsample size. Finally, we compare the resulting confidence intervals with those based on the conventional bootstrap and the Normal approximation method.

5.1. Data and model

We consider the continuously compounded daily returns (expressed in %) for the stock SBM Offshore (SBMO), listed on the NYSE Euronext Amsterdam. We calculate these returns from the stock prices during the period January 1, 2007–April 24,

Table 3

Simulated coverage rates: residual subsample bootstrap (Model 2).

n	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
$\ell = \ell_1$								
500	0.91	0.88	0.89	0.88	0.89	0.88	0.88	0.87
1000	0.89	0.88	0.88	0.88	0.89	0.88	0.89	0.88
5000	0.90	0.89	0.89	0.89	0.89	0.87	0.86	0.88
$\ell = \ell_2$								
500	0.88	0.86	0.86	0.86	0.89	0.88	0.86	0.85
1000	0.88	0.88	0.87	0.89	0.86	0.87	0.85	0.87
5000	0.92	0.90	0.88	0.88	0.89	0.88	0.85	0.86
$\ell = \ell_3$								
500	0.83	0.86	0.84	0.84	0.86	0.86	0.85	0.83
1000	0.87	0.85	0.86	0.85	0.85	0.88	0.84	0.84
5000	0.90	0.90	0.86	0.89	0.88	0.88	0.85	0.85
$\ell = \ell_4$								
500	0.82	0.86	0.82	0.84	0.82	0.85	0.83	0.80
1000	0.87	0.86	0.84	0.85	0.85	0.85	0.83	0.82
5000	0.89	0.90	0.85	0.85	0.88	0.86	0.84	0.83

Notes: For different sample sizes (n), subsample sizes (ℓ) and distributions of the ARMA(1, 1)–GARCH(1, 1) errors z_t (see below), this table displays the simulated coverage rates for the 90% confidence intervals based on the residual subsample bootstrap (based on $B = 1000$ bootstrap replications). The VaR quantile is based on $\varepsilon = 5\%$ (i.e., 5% VaR). The underlying ARMA(1, 1)–GARCH(1, 1) model has parameters $\lambda = (0, 0.5, 0.3, 0.95 \times 20^2/252, 0.4, 0.55)$. For $n = 500$ the subsample sizes equal $\ell_1 = 56$, $\ell_2 = 78$, $\ell_3 = 102$ and $\ell_4 = 156$, respectively. For $n = 1000$ they are 79, 111, 155 and 200. Finally, for $n = 5000$ they are equal to 177, 247, 416 and 585.

Case 1: $z_t \sim N(0, 1)$; Case 2: z_t skewed Normal, with skewness parameter $\xi = 5$ (positive skewness); Case 3: z_t^2 in the domain of attraction of a stable distribution with parameter $a = 1.5$; Case 4: z_t with a symmetric Pareto distribution; Case 5: z_t with a Student's t distribution with 4 degrees of freedom; Case 6: z_t with a skewed Student's t distribution with 4 degrees of freedom and skewness parameter $\xi = 5$ (positive skewness); Case 7: z_t equal to a scaled version of $\exp(0.5L_t)$, where L_t is a Laplace distribution with parameter $b = 2/3$ (positive skewness); Case 8: z_t with a Pareto-type distribution (positive skewness). All distributions have been scaled such that z_t has mean zero and unit variance. The distributions of z_t in Cases 1 and 2 have a finite fourth moment. In the other cases z_t lacks a finite fourth moment. Cases 1, 3, 4, and 5 correspond with symmetric error distributions. Cases 2, 6, 7, and 8 apply to skewed distributions.

Table 4

Simulated coverage rates: residual subsample bootstrap (Model 3).

n	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
$\ell = \ell_1$								
500	0.90	0.87	0.89	0.89	0.88	0.86	0.87	0.78
1000	0.90	0.91	0.87	0.88	0.88	0.87	0.89	0.84
5000	0.89	0.89	0.87	0.88	0.89	0.88	0.88	0.89
$\ell = \ell_2$								
500	0.87	0.89	0.89	0.87	0.88	0.84	0.85	0.77
1000	0.86	0.89	0.86	0.85	0.88	0.87	0.86	0.84
5000	0.90	0.89	0.87	0.87	0.88	0.88	0.86	0.85
$\ell = \ell_3$								
500	0.86	0.90	0.82	0.83	0.85	0.83	0.85	0.73
1000	0.85	0.89	0.86	0.85	0.84	0.85	0.86	0.78
5000	0.90	0.88	0.87	0.86	0.85	0.86	0.88	0.86
$\ell = \ell_4$								
500	0.85	0.86	0.80	0.79	0.81	0.82	0.84	0.70
1000	0.84	0.88	0.83	0.81	0.82	0.81	0.86	0.75
5000	0.88	0.88	0.86	0.84	0.86	0.87	0.86	0.84

Notes: For different sample sizes (n), subsample sizes (ℓ) and distributions of the ARMA(1, 1)–GARCH(1, 1) errors z_t , this table displays the simulated coverage rates for the 90% confidence intervals based on the residual subsample bootstrap (based on $B = 1000$ bootstrap replications). The estimated VaR quantile is based on $\varepsilon = 99\%$ (i.e., 99% VaR). The underlying ARMA(1, 1)–GARCH(1, 1) model has parameters $\lambda = (0, 0.5, 0.3, 0.95 \times 20^2/252, 0.15, 0.8)$. For $n = 500$ the subsample sizes equal $\ell_1 = 56$, $\ell_2 = 78$, $\ell_3 = 102$ and $\ell_4 = 156$, respectively. For $n = 1000$ they are 79, 111, 155 and 200. Finally, for $n = 5000$ they are equal to 177, 247, 416 and 585.

Case 1: $z_t \sim N(0, 1)$; Case 2: z_t skewed Normal, with skewness parameter $\xi = -5$ (negative skewness); Case 3: z_t^2 in the domain of attraction of a stable distribution with parameter $a = 1.5$; Case 4: z_t with a symmetric Pareto distribution; Case 5: z_t with a Student's t distribution with 3 degrees of freedom; Case 6: z_t with a skewed Student's t distribution with 3 degrees of freedom and skewness parameter $\xi = -5$ (negative skewness); Case 7: z_t equal to a scaled version of $-\exp(0.5L_t)$, where L_t is a Laplace distribution with parameter $b = 2/3$ (positive skewness); Case 8: z_t with a Pareto-type distribution (negative skewness). All distributions have been scaled such that z_t has mean zero and unit variance. The distributions of z_t in Cases 1 and 2 have a finite fourth moment. In the other cases z_t lacks a finite fourth moment. Cases 1, 3, 4, and 5 correspond with symmetric error distributions. Cases 2, 6, 7, and 8 apply to skewed distributions.

Fig. 1. Raw returns and GARCH residuals. *Notes:* Figure (a) displays the raw returns on the SBMO stock for the full sample period (January 2, 2007–April 24, 2013). Figure (b) shows a histogram of the GARCH(1, 1) residuals corresponding with the first rolling-window period (January 2, 2007–December 30, 2011). Figure (c) depicts the QQ-plot for these residuals.

2013. The data have been downloaded from Datastream, where their mnemonic is H:SBMO. This results in a sample of 1647 daily returns. The presence of GARCH effects is visible from Fig. 1(a), where high (absolute) returns tend to be followed by more high returns and low (absolute) returns tend to be followed by more low returns. This figure illustrates that the SBMO stock has experienced several substantial price drops and recoveries since the subprime crisis in 2007 and the subsequent start of the global financial crisis that was initiated with the fall of Lehman Brothers in September 2008.

At first sight it might seem tempting to leave out or to correct a few extreme observations to get a more well-behaved return series. However, *VaR analysis is the analysis of extremes*. Leaving out or adjusting extreme observations will result in underestimation of extreme future losses. For banks and insurance companies this underestimation might eventually translate into a capital reserve that is too low, with bankruptcy as the most dramatic consequence. We will therefore not leave out or correct any observations. Because extreme observations might distort QML estimation of the ARMA–GARCH model (Carnero et al., 2007), we will explicitly verify that this is not the case.

We apply a rolling-window analysis that starts with the period January 2, 2007–December 30, 2011 and ends with the period April 24, 2008–April 23, 2013. For each of the 343 resulting rolling-windows periods (consisting of $n = 1304$ observations each), we estimate a GARCH(1, 1) model with constant expected returns (i.e., $\mu_t = \mu$) by means of QML. The parameter estimates for the first rolling-window period equal $\hat{\mu} = 0.125$ (standard error 0.055), $\hat{\omega} = 0.183$ (0.061), $\hat{\alpha} = 0.137$ (0.024), and $\hat{\beta} = 0.847$ (0.027). As often found for daily returns, the volatility persistence is close to unity. Extreme observations do not distort QML estimation of the GARCH model; i.e., leaving out the most extreme raw returns does only slightly affect the GARCH(1, 1) model's volatility forecast used to estimate VaR. A histogram and QQ-plot of the corresponding residuals are shown in Figs. 1(b) and (c). The histogram illustrates the negative skewness of the model errors' distribution. The GARCH errors' lack of normality becomes apparent from the Normal QQ-plot, which shows that both tails of the residual distribution are relatively heavy. The results for the other rolling-window periods are similar and omitted to save space.

We perform several statistical tests to verify that the GARCH(1, 1) model with constant expected returns provides a good fit to the returns on the SBMO stock. We apply Ljung–Box tests to test for autocorrelation and remaining GARCH effects. Furthermore, we follow a backtesting procedure to test for correct unconditional coverage of the estimated 1% and 99% VaR. Detailed test results can be found in the appendix with supplementary material (see Appendix D). The diagnostic tests and backtesting results indicate that the choice for a GARCH(1, 1) model is reasonable.

5.2. Replicating histogram

In order to apply the residual subsample bootstrap to real-life data, it is neither necessary to know whether the model errors possess a finite fourth moment nor to know the QML estimator's asymptotic distribution. However, inefficiency is the price that is paid for applying the residual subsample bootstrap in situations that the conventional bootstrap would be valid. To avoid such inefficiencies, we might prefer to apply the residual subsample bootstrap only when the conventional bootstrap is inconsistent. However, the development of a formal selection rule to decide between the residual subsample bootstrap and the conventional bootstrap is beyond the scope of this article and left as a topic for future research.

Here we consider the replicating histogram of Sherman and Carlstein (1996) to get some preliminary insight in the asymptotic behavior of the QML estimator when applied to the SBMO stock returns. This technique employs a (residual) subsample bootstrap to assess the shape of a statistic's sampling distribution. The replicating histogram is strongly consistent under weak regularity conditions, which are virtually the same as the ones described in Section 4. In our setting, the replicating histogram boils down to a histogram (or kernel density estimate) of the residual subsample replicates

Fig. 2. Replicating histograms. *Notes:* The replicating histograms in (a) and (b) apply to the asymptotic distribution of the VaR estimates, as derived from the pure GARCH(1, 1) model estimated over the first rolling-window period (January 2, 2007–December 30, 2011). The R-routine `bkde` in the `KernSmooth` library has been used to create the kernel density plots, with the bandwidth determined by the `dpik` command from the same library.

$\tilde{\xi}_{n|n+1,1}^{\varepsilon}, \dots, \tilde{\xi}_{n|n+1,B}^{\varepsilon}$. Only the shape of the replicating histogram is informative (e.g., symmetry versus skewness). For this reason the axes of the replicating histogram are always left blank. When the model errors possess a finite fourth moment, the QML-based VaR estimator is asymptotically normally distributed (Goa and Song, 2008). Consequently, when the replicating histogram reveals a non-Normal shape, the model errors cannot have a finite fourth moment. Similarly, a replicating histogram with a non-Normal shape precludes the situation that the model errors lack a finite fourth moment while the squared model errors are in the domain of attraction of the Normal distribution (Hall and Yao, 2003). Figs. 2(a) and (b) show kernel-smoothed versions of the replicating histogram for the 1% and 99% VaR estimates based on the first rolling-window period (for $B = 10\,000$ and $\ell_3 = 183$), which reveals leptokurtic and skewed distributions and casts doubts on the model errors' finite fourth moment. Similar non-Normal shapes are found for the other rolling-window samples. These results are robust to the choice of the subsample size.

5.3. Confidence intervals

Figs. 3(a)–(b) display the 1% and 99% VaR estimates for the SBMO stock returns, together with the raw returns. The 1% VaR ranges between -16.0% and -4.3% over the rolling window horizon, whereas the 99% VaR takes values between 3.7% and 13.3% . Figs. 3(a)–(b) also depict the estimated 90% confidence intervals, based on the residual subsample bootstrap. The residual subsample bootstrap is based on $B = 10\,000$ bootstrap replications and subsample size $\ell_3 = 183$. Additional bootstrapping shows that the empirical CPEs are relatively robust to the choice of the subsample size. The chosen subsample size $\ell_3 = 183$ ensures that the empirical CPEs (as defined in Section 4.1) remain within the chosen threshold of $\delta = 0.10$.

For risk management purposes, it is important to realize that the VaR estimates' distribution is subject to substantial estimation uncertainty as shown in Figs. 3(a)–(b). As a stylized example, suppose that the 1% VaR has been estimated by an insurance company in order to set a capital reserve. In such a situation additional capital on top of the VaR estimate will be required to account for the underlying estimation uncertainty. For example, instead of using the VaR point estimate, one could consider using the confidence interval's lower bound to set a reserve.

For the sake of comparison we also obtain confidence intervals for the 1% and 99% VaR using the conventional bootstrap (with $B = 1000$) and the Normal approximation method (with tail sample size $k = \lfloor 1.5(\log n)^2 \rfloor = 77$). Figs. 4(a)–(c) show the rolling-window point estimates of the one-day ahead 99% VaR, together with 90% point-wise confidence intervals based on the conventional bootstrap, the Normal approximation method and the residual subsample bootstrap. Again the raw returns on the SBMO stock are also displayed in the figures. The confidence intervals based on the Normal approximation are slightly wider on average than those derived from the conventional bootstrap; their average radii are 1.27 and 1.06, respectively. With an average radius of 3.12 the confidence intervals based on the residual subsample bootstrap are on average 2.5 to 3 times as wide as those obtained by means of the Normal approximation method and the conventional bootstrap.

Which of the three estimated confidence intervals is to be preferred in this empirical example? When the QML estimator's asymptotic distribution is Normal, the residual subsample bootstrap's confidence intervals are unnecessarily wide. In such a scenario it would be more efficient to apply the conventional bootstrap or to derive analytical confidence intervals in the spirit of Goa and Song (2008). However, the situation is different when the QML estimator's asymptotic distribution is not Normal, in which case the conventional bootstrap turns out inconsistent. In particular, our simulation study has made clear that the confidence intervals based on the conventional bootstrap can turn out too narrow in the presence of skewed model errors without finite fourth moment, with poor coverage rates as a result. In such a scenario also the Normal approximation method fails. Furthermore, we already mentioned the sensitivity of the Normal approximation method to

Fig. 3. Confidence intervals for 1% and 99% VaR: residual subsample bootstrap. *Notes:* In each figure the solid black line shows the rolling-window estimates of the one-day ahead VaR (i.e., $\hat{\xi}_{t|t+1}^e$), where t denotes a date between December 30, 2011 and April 23, 2013. The solid blue line indicates the returns on the SBMO stock at day $t + 1$ (i.e., r_{t+1}), allowing for a comparison between realized returns r_{t+1} and $\hat{\xi}_{t|t+1}^e$. The gray area between the two red lines corresponds with a point-wise 90% confidence interval. The residual subsample bootstrap is based on subsample size $\ell_3 = 183$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Confidence intervals for 99% VaR: three different methods. *Notes:* In each figure the solid black line shows the rolling-window estimates of the one-day ahead VaR (i.e., $\hat{\xi}_{t|t+1}^e$), where t denotes a date between December 30, 2011 and April 23, 2013. The solid blue line indicates the returns on the SBMO stock at day $t + 1$ (i.e., r_{t+1}), allowing for a comparison between realized returns r_{t+1} and $\hat{\xi}_{t|t+1}^e$. The gray area between the two red lines corresponds with a point-wise 90% confidence interval. The Normal approximation method uses tail sample size $k = \lfloor 1.5(\log n)^2 \rfloor = 77$, while the residual subsample bootstrap is based on subsample size $\ell_3 = 183$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the choice of the tail sample size. In sum, the worst-case scenario for the confidence intervals based on the conventional bootstrap and the Normal approximation method is very poor coverage, while unnecessarily wide confidence intervals is the worst-case scenario for the residual subsample bootstrap. Given the replicating histogram's evidence against the QML estimator's asymptotic normality, we are on the safe side when we opt for the wider confidence intervals based on the residual subsample bootstrap.

6. Conclusions

When the ARMA–GARCH model errors lack a finite fourth moment, the asymptotic distribution of the QML estimator may not be Normal. In such a case the conventional bootstrap turns out inconsistent. Surprisingly, simulations show that the conventional bootstrap, despite its inconsistency, provides accurate confidence intervals for ARMA–GARCH VaR in case of symmetric error distributions without finite fourth moment. The bootstrap does fail, however, in the presence of skewed error distributions without finite fourth moment. In this case also the Normal approximation and data tilting methods proposed by Chan et al. (2007) fail.

We propose a residual subsample bootstrap to obtain confidence intervals for ARMA–GARCH VaR. According to theory, this ‘omnibus’ method generates confidence intervals with asymptotically correct coverage rates under very mild conditions. Simulations illustrate the residual subsample bootstrap's favorable finite-sample properties. It produces confidence intervals with good coverage rates for a wide range of specifications, while being relatively robust to the choice of the subsample size.

The residual subsample bootstrap even performs satisfactorily when other methods fail in the presence of skewed model errors without finite fourth moment.

The price that is paid for the omnibus nature of the subsample bootstrap is inefficiency. Because the speed of convergence and the limiting distribution of the QML-based VaR estimator are typically unknown when the ARMA–GARCH model errors lack a finite fourth moment, it is nevertheless attractive to use the subsample bootstrap for confidence interval estimation despite its inefficiency.

The approach proposed in this study can be extended to multi-period VaR estimates and other risk measures such as expected shortfall. We leave this as a topic for future research.

Acknowledgments

The author is grateful to Edward Carlstein, Ruud Koning, Michael Sherman, Tom Wansbeek, the participants of the ESEM 2013 conference in Göteborg, as well as two anonymous referees and the guest editor for useful comments and suggestions. She also gratefully acknowledges financial support by a Vidi grant (452.11.007) in the ‘Vernieuwingsimpuls’ program of the Netherlands Organization for Scientific Research (NWO). The usual disclaimer applies.

Appendix A. Conventional residual bootstrap for ARMA–GARCH models

1. Fit an ARMA–GARCH model to the observed sample $r_1, \dots, r_n$ by means of QML, resulting in $\hat{\lambda}$ (the vector of estimated ARMA–GARCH parameters), $\hat{Q}_z^\varepsilon$, $\hat{\mu}_t$ and $\hat{\sigma}_t$ for $t = 1, \dots, n$.
2. Draw a random sample of length n with replacement from the ARMA–GARCH model residuals $\hat{z}_1, \dots, \hat{z}_n$, yielding $\tilde{z}_1, \dots, \tilde{z}_n$.
3. Use $\hat{\lambda}$ and $\tilde{z}_1, \dots, \tilde{z}_n$ to recursively generate values $\tilde{r}_1, \dots, \tilde{r}_n$ according to Eqs. (1)–(3).
4. Use the bootstrap sample $\tilde{r}_1, \dots, \tilde{r}_n$ to estimate the ARMA–GARCH model by means of QML, yielding $\tilde{\lambda}$, $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ for $t = 1, \dots, n$.
5. Define Q_z^ε as the $\varepsilon\%$ sample quantile of $\tilde{z}_t = (\tilde{r}_t - \tilde{\mu}_t)/\tilde{\sigma}_t$, $t = 1, \dots, n$ and use it as a bootstrap estimator of Q_z^ε .
6. Use $\tilde{\lambda}$ and the observed sample $r_1, \dots, r_n$ to statically calculate $\tilde{\mu}_{t+1}^*$ and $\tilde{\sigma}_{t+1}^*$ for $t = 1, \dots, n$ according to Eqs. (2)–(3).
7. Calculate $\hat{\xi}_{n|n+1}^\varepsilon$ from Eq. (5), using $\tilde{\mu}_{n+1}^*$, $\tilde{\sigma}_{n+1}^*$, and $\hat{Q}_z^\varepsilon$.

Steps 1–7 are repeated B times, resulting in $\hat{\xi}_{n|n+1,1}^\varepsilon, \dots, \hat{\xi}_{n|n+1,B}^\varepsilon$. The proposed $x\%$ confidence interval follows directly from the *percentile method* (see e.g. Efron and Tibshirani, 1993) and has the form

$$\left[\hat{\xi}_{n|n+1,1}^\varepsilon + Q_\xi^{(1-x)/2}, \hat{\xi}_{n|n+1,1}^\varepsilon + Q_\xi^{(1+x)/2} \right]. \quad (\text{A.1})$$

Here Q_ξ^y is defined as the empirical $y\%$ quantile of $\hat{\xi}_{n|n+1,1}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon, \dots, \hat{\xi}_{n|n+1,B}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon$, with $\hat{\xi}_{n|n+1}^\varepsilon$ calculated from $\hat{\mu}_{n+1}$, $\hat{\sigma}_{n+1}$, and $\hat{Q}_z^\varepsilon$ according to Eq. (5).

Hall and Yao (2003) prove that non-normality of a GARCH model’s QML estimator arises when the fourth moment of the model errors is not finite and the distribution of the squared model errors is not in the domain of attraction of the Normal distribution. Subsequently, they show that the conventional bootstrap turns out inconsistent in such a scenario.

Appendix B. Bootstrapping conditional quantiles in ARMA–GARCH models

We are interested in the conditional quantile $\xi_{n+1}^\varepsilon = \mu_{n+1} + \sigma_{n+1}Q_z^\varepsilon$. We can write $\mu_{n+1} = \mu_{n+1}(r_1, \dots, r_n; \lambda)$ and $\sigma_{n+1} = \sigma_{n+1}(r_1, \dots, r_n; \lambda)$, where λ denotes the vector of ARMA–GARCH parameters and $r_1, \dots, r_n$ the observed sample. Hence, the conditional quantile $\xi_{n|n+1}^\varepsilon$ boils down to a function of λ and Q_z^ε (the fixed parameters), as well as $r_1, \dots, r_n$ (the observed sample). With $Q = Q_z^\varepsilon$ for simplicity, we can thus write $\xi_{n|n+1}^\varepsilon = R(r_1, \dots, r_n; \lambda, Q)$, for some function $R(\cdot)$.

Many theoretical studies about the bootstrap consider a fixed target parameter as the object of interest. Our case looks a bit different because we are interested in bootstrapping $R(r_1, \dots, r_n; \lambda, Q)$. Here $R(r_1, \dots, r_n; \lambda, Q)$ represents a random variable that does not only depend on the fixed parameters λ and Q , but also on the *observed sample* $r_1, \dots, r_n$. This gives rise to the question whether the results in Sherman and Carlstein (2004) still apply to our case where the object of interest is $R(r_1, \dots, r_n; \lambda, Q)$. That is, the results in Sherman and Carlstein (2004) apply at first sight only to the case of a fixed parameter.

Let us first consider this issue in the context of the traditional (n -out-of- n with-replacement) bootstrap. The validity of the traditional bootstrap extends to the case where the quantity of interest is not a fixed parameter (say) θ but a random variable of the form $R(X_1, \dots, X_n; \theta)$, where $X_1, \dots, X_n$ are the observed data. This is shown in e.g. Efron (1982, p. 43) and Efron and Gong (1983, pp. 42–43).

It is useful to give some background here. Random variables of the form $R(X_1, \dots, X_n; \theta)$ have a long history in the bootstrap literature. Traditionally, such random variables are called ‘roots’; see e.g. Politis et al. (1999, p. 4). This explains why the letter R is generally used to denote such functions. The use of the word ‘root’ stems from the fact that the most well-known root studied in the bootstrap literature is $R(X_1, \dots, X_n; \theta) = \hat{\theta} - \theta$; i.e., the bias function. This function is a true ‘root’

in the sense that one is interested in the value of θ for which $\mathbb{E}(R(X_1, \dots, X_n; \theta)) = 0$. However, also more general functions of the form $R(X_1, \dots, X_n; \theta)$ are referred to as 'roots' to emphasize their dependence on the observed sample instead of a fixed parameter only. As shown by Efron (1982) and Efron and Gong (1983), the bootstrap can be safely applied to both fixed parameters and roots.

Now we turn to the specific case of the subsample bootstrap of Sherman and Carlstein (2004). This bootstrap is valid under the conditions mentioned in their study, which are the same conditions as mentioned in Bertail et al. (1999). Neither Sherman and Carlstein (2004) nor Bertail et al. (1999) mention the extension of their subsample bootstrap from fixed parameters to random variables of the form $R(X_1, \dots, X_n; \theta)$, but Politis et al. (1999, Section 2.5) do so in their book about subsampling. They make clear that the validity of the subsample bootstrap methods of Bertail et al. (1999) and Sherman and Carlstein (2004) extends to roots, just like the traditional bootstrap does. This also becomes apparent from their proofs showing the theoretical validity of subsampling, in which it does not matter whether the object of interest is a fixed parameter or a root.

Appendix C. Residual subsample bootstrap for ARMA–GARCH models

1. Fit an ARMA–GARCH model to the observed sample $r_1, \dots, r_n$ by means of QML, resulting in $\hat{\lambda}$ (the vector of estimated ARMA–GARCH parameters), $\hat{Q}_z^\varepsilon$, $\hat{\mu}_t$ and $\hat{\sigma}_t$ for $t = 1, \dots, n$.
2. Draw a random subsample of length ℓ without replacement from the ARMA–GARCH model residuals $\hat{z}_1, \dots, \hat{z}_n$, yielding $\tilde{z}_1, \dots, \tilde{z}_\ell$.
3. Use $\hat{\lambda}$ and $\tilde{z}_1, \dots, \tilde{z}_\ell$ to recursively generate values $\tilde{r}_1, \dots, \tilde{r}_\ell$ according to Eqs. (1)–(3).
4. Use the bootstrap sample $\tilde{r}_1, \dots, \tilde{r}_\ell$ to estimate the ARMA–GARCH model by means of QML, yielding $\tilde{\lambda}$, $\tilde{\mu}_t$ and $\tilde{\sigma}_t$ for $t = 1, \dots, \ell$.
5. Define $\tilde{Q}_z^\varepsilon$ as the $\varepsilon\%$ sample quantile of $\tilde{z}_t = (\tilde{r}_t - \tilde{\mu}_t)/\tilde{\sigma}_t$, $t = 1, \dots, \ell$ and use it as a bootstrap estimator of Q_z^ε .
6. Use $\tilde{\lambda}$ and the observed sample $r_1, \dots, r_n$ to statically calculate $\tilde{\mu}_{t+1}^*$ and $\tilde{\sigma}_{t+1}^*$ for $t = 1, \dots, n$ according to Eqs. (2)–(3).
7. Calculate $\tilde{\xi}_{n|n+1}^\varepsilon$ from Eq. (5), using $\tilde{\mu}_{n+1}^*$, $\tilde{\sigma}_{n+1}^*$, and $\tilde{Q}_z^\varepsilon$.

Steps 1–7 are repeated B times, resulting in $\tilde{\xi}_{n|n+1,1}^\varepsilon, \dots, \tilde{\xi}_{n|n+1,B}^\varepsilon$. The proposed $x\%$ confidence interval has the form

$$\left[\tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1+x)/2}, \tilde{\xi}_{n|n+1,1}^\varepsilon - Q_\xi^{(1-x)/2} \right]. \quad (\text{C.1})$$

Here Q_ξ^y is defined as the empirical $y\%$ quantile of $\tilde{\xi}_{n|n+1,1}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon, \dots, \tilde{\xi}_{n|n+1,B}^\varepsilon - \hat{\xi}_{n|n+1}^\varepsilon$, with $\hat{\xi}_{n|n+1}^\varepsilon$ calculated from $\hat{\mu}_{n+1}$, $\hat{\sigma}_{n+1}$, and $\hat{Q}_z^\varepsilon$ according to Eq. (5).

Remarks.

- Because it is not possible to use all $\binom{n}{\ell}$ subsamples of size ℓ from the n residuals, we follow Politis and Romano (1994) and randomly draw $B = 1000$ of such subsamples without replacement. This choice of B turns out adequate, in the sense that a further increase in B has only a minor impact on the confidence intervals.
- As noted by Sherman and Carlstein (2004), the choice of $\tilde{\xi}_{n|n+1,i}^\varepsilon$ is arbitrary and can be replaced by any $\tilde{\xi}_{n|n+1,i}^\varepsilon$, $i = 2, \dots, B$.

Appendix D. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.csda.2014.08.011>.

References

- Athreya, K., 1987a. Bootstrap of the mean in the infinite variance case. In: Prohorov, Y., Sazonov, V. (Eds.), Proceedings of the First World Congress of the Bernoulli Society, vol. 5. VNU Science Press, Utrecht, The Netherlands, pp. 95–98.
- Athreya, K., 1987b. Bootstrap of the mean in the infinite variance case. *Ann. Statist.* 15, 724–731.
- Bams, D., Lehnert, T., Wolff, C., 2005. An evaluation framework for alternative VaR-models. *J. Int. Money Finance* 10, 944–958.
- Bertail, P., Politis, D., Romano, J., 1999. On subsampling estimators with unknown rate of convergence. *J. Amer. Statist. Assoc.* 94, 569–579.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31, 307–327.
- Bollerslev, T., Wooldridge, J., 1992. Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Rev.* 11, 143–172.
- Carnero, M., Penã, D., Ruiz, E., 2007. Effects of outliers on the identification and estimation of GARCH models. *J. Time Ser. Anal.* 28, 471–497.
- Chan, N., Deng, S.-J., Peng, L., Xia, Z., 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *J. Econometrics* 137, 556–576.
- Chen, S., Tang, C., 2005. Nonparametric inference of value-at-risk for dependent financial returns. *J. Financ. Econ.* 3, 227–255.
- Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. *J. Risk* 7, 1–28.
- Dowd, K., 2002. *Measuring Market Risk*. Wiley, Chichester, UK.
- Duan, J., 1994. Maximum likelihood estimation using price data of the derivative contract. *Math. Finance* 4, 155–167.
- Efron, B., 1982. The jackknife, the bootstrap and Other Resampling Plans. SIAM, Philadelphia.
- Efron, B., Gong, G., 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Statist.* 37, 26–48.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. In: Monographs on Statistics and Applied Probability, vol. 57. Chapman & Hall, New York.

- Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R., Manganelli, S., 2004. CAViaR: conditional autoregressive value-at-risk by regression quantiles. *J. Bus. Econom. Statist.* 22, 367–381.
- Francq, C., Zakoian, J.-M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C., Zakoian, J.-M., 2012. Risk-parameter estimation in volatility models. MPRA Working Paper.
- Goa, F., Song, F., 2008. Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24, 1404–1424.
- Grabchak, M., Samorodnitsky, D., 2010. Do financial returns have finite or infinite variance? A paradox and an explanation. *Quant. Finance* 10, 883–893.
- Hall, P., 1990. Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab.* 18, 1342–1360.
- Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer Series in Statistics, New York.
- Hall, P., Yao, Q., 2003. Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
- Hill, B., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.* 3, 1163–1174.
- Horowitz, J., 2001. The bootstrap. In: Heckman, J., Leamer, E. (Eds.), *Handbook of Econometrics*, vol. 5. Elsevier, Amsterdam, pp. 3159–3228.
- Huang, D., Wang, H., Yao, Q., 2008. Estimating GARCH models: when to use what? *Econom. J.* 11, 27–38.
- Huisman, R., Koedijk, K., Kool, C., Palm, F., 2001. Tail-index estimates in small samples. *J. Bus. Econom. Statist.* 19, 208–216.
- Jorion, P., 1996. Risk²: measuring their risk in value-at-risk. *Financ. Anal. J.* 52, 47–56.
- Kinader, J., 1992. An invariance principle applicable to the bootstrap. In: Lepage, R., Billard, L. (Eds.), *Exploring the Limits of Bootstrap*. In: Wiley Series in Probability and Mathematical Statistics, pp. 157–181.
- Knight, K., 1989. On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* 17, 1168–1175.
- Li, W., Ling, S., McAleer, M., 2002. Recent theoretical results for time series models with GARCH errors. *J. Econ. Surv.* 16, 245–269.
- Luger, R., 2012. Finite-sample bootstrap inference in GARCH models with heavy-tailed innovations. *Comput. Statist. Data Anal.* 56, 3198–3211.
- Mammen, E., 1992. When Does Bootstrap Work? Asymptotic Results and Simulations. In: *Lecture Notes in Statistics*, vol. 77. Springer, New York.
- McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empir. Finance* 7, 271–300.
- Papadimitis, E., Politis, D., 2002. The tapered block bootstrap for general statistics from stationary sequences. *Econom. J.* 5, 131–148.
- Pascal, L., Romo, J., Ruiz, E., 2006. Bootstrap prediction for returns and volatilities in GARCH models. *Comput. Statist. Data Anal.* 50, 2293–2312.
- Peng, L., Qi, L., 2006. Confidence regions for high quantiles of a heavy tailed distribution. *Ann. Statist.* 34, 1964–1986.
- Peng, L., Yao, Q., 2003. Least absolute deviations estimator for ARCH and GARCH models. *Biometrika* 90, 967–975.
- Politis, D., Romano, J., 1994. Large-sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* 22, 2031–2050.
- Politis, D., Romano, J., Wolf, M., 1999. *Subsampling*. Springer Series in Statistics, New York.
- Sandström, A., 2011. *Handbook of Solvency for Actuaries and Risk Managers: Theory and Practice*. Chapman & Hall/CRC Finance, Boca Raton.
- Sen, B., Banerjee, M., Woodroffe, M., 2010. Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* 38, 1953–1977.
- Sherman, M., Carlstein, E., 1996. Replicating histograms. *J. Amer. Statist. Assoc.* 91, 566–576.
- Sherman, M., Carlstein, E., 2004. Confidence intervals based on estimators with unknown rates of convergence. *Comput. Statist. Data Anal.* 46, 123–139.
- Xiao, Z., Koenker, R., 2006. Quantile autoregression. *J. Amer. Statist. Assoc.* 101, 980–990.