Estimation Risk in Financial Risk Management*

Peter Christoffersen
McGill University and CIRANO

Sílvia Gonçalves
Université de Montréal and CIRANO

December 30, 2004

Abstract

Value-at-Risk (VaR) is increasingly used in portfolio risk measurement, risk capital allocation and performance attribution. Financial risk managers are therefore rightfully concerned with the precision of typical VaR techniques. The purpose of this paper is to assess the precision of common dynamic models and to quantify the magnitude of the estimation error by constructing confidence intervals around the point VaR and expected shortfall (ES) forecasts. A key challenge in constructing proper confidence intervals arises from the conditional variance dynamics typically found in speculative returns. Our paper suggests a resampling technique which accounts for parameter estimation error in dynamic models of portfolio variance.

JEL Number: G12.

Keywords: Risk management, bootstrapping, GARCH

*Christoffersen is at the Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, Quebec, Canada H3A 1G5, Phone: (514) 398-2969, Fax: (514) 398-3876, Email: peter.christoffersen@mcgill.ca, Web: www.christoffersen.ca. Gonçalves is at Département de sciences économiques, Université de Montréal, C.P.6128, succ. Centre-Ville, Montréal, Québec, Canada H3C 3J7, Phone: (514)343-6556, Fax: (514)343-7221, Email: silvia.goncalves@umontreal.ca, Web: http://mapageweb.umontreal.ca/goncals. We are grateful for comments particulary from the Editor Philippe Jorion as well as from Sean Campbell, Valentina Corradi, Frank Diebold, Jin Duan, René Garcia, Éric Jacquier, Simone Manganelli, Stefan Mittnik, Nour Meddahi, Matt Pritsker, Éric Renault, and Enrique Sentana. FQRSC, IFM2 and SSHRC provided financial support. The usual disclaimer applies.

1 Motivation

Value-at-Risk (VaR) is increasingly used in portfolio risk measurement, risk capital allocation and performance attribution, and financial risk managers are rightfully concerned with the precision of typical VaR techniques. VaR is defined as the conditional quantile of the portfolio loss distribution for a given horizon (typically a day or a week) and for a given coverage rate (typically 1% or 5%), and the expected shortfall (ES) is defined as the expected loss beyond the VaR. The VaR and ES measures are thus statements about the left tail of the return distribution and in realistic sample sizes (500 or 1000 daily observations) such statements are likely to be made with considerable error.

The purpose of this paper is twofold: First, we want to assess the potential loss of accuracy from estimation error when calculating VaR and ES. Second, we want to assess our ability to quantify ex-ante the magnitude of this error via the construction of confidence intervals around the VaR and ES measures. This issue of estimation risk for VaR has been considered previously in the i.i.d. return case by for example Jorion (1996) and Pritsker (1997). But a key challenge in constructing proper VaR and ES confidence intervals arises from the conditional variance dynamics typically found in speculative returns. We quantify these dynamics using the celebrated GARCH model of Engle (1982) and Bollerslev (1986). Due to its ability to capture salient features of the return dynamics in very parsimonious and easily estimated specifications, GARCH has become the workhorse model in financial risk management. Nevertheless, and surprisingly, very little is known about the uncertainty in the GARCH VaR and ES forecasts arising from parameter estimation error.[1]

Our paper extends the resampling technique of Pascual, Romo and Ruiz (2001), which accounts for parameter estimation error in dynamic models of portfolio variance, to the case of VaR and ES forecasts. To our knowledge no asymptotic theory has been established for calculating confidence intervals for risk measures in this context. The resampling technique we propose can be relatively easily extended to longer horizons, to multivariate risk models, and to allowing for model specification error.

Our Monte Carlo evidence suggests that when assuming independent returns the bootstrap intervals work well for the commonly used Historical Simulation VaR model. However, when allowing for realistic GARCH effects the Historical Simulation VaR implies nominal 90% confidence intervals for the one-day, 1% VaR that are much too narrow. Historical Simulation essentially

[1] Baillie and Bollerslev (1992) construct approximate prediction intervals for GARCH variance forecasts at multiple horizons but ignore estimation error. Furthermore, risk management suveys and textbooks such as for example Christoffersen (2003), Duffie and Pan (1997), and Jorion (2000) give little or no attention to the estimation error issue.

ignores the time varying risk from GARCH and the finding of poor confidence intervals is therefore not surprising in this case. Methods which properly account for conditional variance dynamics, such as Filtered Historical Simulation (FHS) suggested by Hull and White (1998) and Barone-Adesi et al (1998, 1999), imply 90% VaR confidence intervals that contain close to 90% of the true VaRs.

In our benchmark GARCH case, the average width of the VaR interval for the best model is 27-38% of the true VaR depending on the estimation sample size. The average width of the ES confidence interval is 22-42% of the true ES value (again depending on the sample size) for the best model. Estimation risk is thus found to be substantial even in tightly parameterized models. Importantly, we find that it is in general more difficult to construct accurate confidence intervals for the ES measure. Typically, the confidence intervals from risk models we consider tend to contain the true ES less frequently than the 90% they should.

Accurate confidence intervals reported along with the VaR point estimate will facilitate the use of VaR in active portfolio management as the following example illustrates: Consider a portfolio manager who is allowed to take on portfolios with a VaR of up to 15% of the current capital. If the risk manager calculates the actual point estimate VaR to be 13% with a confidence interval of 10-16% then the cautious portfolio manager should rebalance the portfolio to reduce risk because the 16% confidence interval upper limit is above the VaR limit. Relying instead only on the point estimate of 13% would not signal any need to rebalance.

The remainder of the paper is organized as follows. Section 2 presents our conditionally nonnormal GARCH portfolio return generating process and defines five risk models which we will consider in the subsequent analysis. Section 3 presents the resampling methods used to generate the VaR and ES confidence intervals. Section 4 presents the Monte Carlo setup and discusses the results we obtained. Finally, Section 5 concludes and suggests avenues for future research.

2 Model and Risk Measures

In this paper we model the dynamics of the daily losses (the negative of returns) on a given financial asset or portfolio according to the model

$$L_t = \sigma_t \varepsilon_t, \quad t = 1, \dots, T, \tag{1}$$

where ε_t are i.i.d. with mean zero, variance one, and distribution function G. In particular, we focus on the case in which G is a standardized Student's t distribution with d degrees of freedom,[2]

[2] The model can be generalized to allow for skewness following Theodossiou (1998). See also, Tsay (2002) Chapter 7.

i.e.

$$\sqrt{d/(d-2)}\varepsilon_t \sim t(d).$$

To model the volatility dynamics we use a symmetric GARCH(1,1) model for σ_t^2:

$$\sigma_t^2 = \omega + \alpha L_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where $\alpha+\beta < 1$. The GARCH(1,1) model with standardized Student's t innovations has been very successful in capturing the volatility clustering and nonnormality found in daily asset return data. See for example Bollerslev (1987) and Baillie and Bollerslev (1989). Although we focus on this particular model of returns, our approach applies to more complex models of σ_t^2 and/or to other distributions for ε_t.

At a given point in time, we are interested in describing the risk in the tails of the conditional distribution of losses over a given horizon, say one-day, using all the information available at that time. We consider two popular risk measures. One is the Value-at-Risk (VaR), which is simply a conditional quantile of the losses distribution. The other is the Expected Shortfall (ES), which measures the expected losses over the next day given that losses exceed VaR.

The VaR measure for time $T+1$ with coverage probability p, based on information at time T, is defined as the (positive) value VaR_{T+1}^p such that

$$\Pr\left(L_{T+1} > VaR_{T+1}^p | \mathcal{F}^T\right) = p, \tag{2}$$

where $\mathcal{F}^T$ denotes the information available at time T. Typically p is a small number, e.g. $p = 0.01$ or $p = 0.05$.

Similarly, we define the ES measure for time $T+1$ with coverage probability p, given information at time T, as the (positive) value ES_{T+1}^p such that

$$ES_{T+1}^p = E\left(L_{T+1} | L_{T+1} > VaR_{T+1}^p, \mathcal{F}^T\right). \tag{3}$$

Given model (1), we can obtain simplified expressions for VaR_{T+1}^p and ES_{T+1}^p. More specifically, we can show that

$$VaR_{T+1}^p = \sigma_{T+1} G_{1-p}^{-1} \equiv \sigma_{T+1} c_{1,p}, \tag{4}$$

where G_{1-p}^{-1} denotes the $(1-p)$-th quantile of G, the distribution of standardized losses $\varepsilon_t = L_t/\sigma_t$, and σ_{T+1} is the conditional volatility for time $T+1$. For instance, if G is the standard normal distribution Φ and $p = 0.05$, we have that $G_{1-p}^{-1} = \Phi_{0.95}^{-1} = 1.645$, and thus $VaR_{T+1}^p = 1.645\sigma_{T+1}$. In the general case where $\varepsilon \sim G$, equation (4) shows that we can express VaR_{T+1}^p as the product of σ_{T+1} with a constant $c_{1,p} \equiv G_{1-p}^{-1}$, whose value depends on G and on p.

Similarly, under model (1), we can show that

$$ES^p_{T+1} = \sigma_{T+1} E\left(\varepsilon | \varepsilon > G^{-1}_{1-p}\right) \equiv \sigma_{T+1} c_{2,p}, \tag{5}$$

where ε is an i.i.d. random variable with mean zero, variance one, and distribution G. If $\varepsilon \sim N(0,1)$, we can show that $E(\varepsilon|\varepsilon > a) = \frac{\phi(a)}{1-\Phi(a)}$, for any constant a, where ϕ and Φ denote the density and the distribution functions of a standard normal random variable. Thus, in this particular case, $ES^p_{T+1} = \sigma_{T+1}\frac{\phi\left(\Phi^{-1}_{1-p}\right)}{p}$ and $c_{2,p} \equiv \frac{\phi\left(\Phi^{-1}_{1-p}\right)}{p}$. When ε has a standardized Student distribution with d degrees of freedom, $c_{2,p}$ is given by a different formula. To describe this formula, let t_d be a random variable following a Student-t distribution with d degrees of freedom. Andreev and Kanto (2004) show that for any a, $E(t_d|t_d > a) = \left(1 + \frac{a^2}{d}\right)\frac{d}{d-1}\frac{f(a)}{1-F(a)}$, where f and F denote the probability density and the cumulative density functions of t_d. Using this result, we can show that in this case,

$$c_{2,p} \equiv E\left(\varepsilon|\varepsilon > G^{-1}_{1-p}\right) = \left(1 + \left(\sqrt{\frac{d}{d-2}}G^{-1}_{1-p}\right)^2/d\right)\frac{d}{d-1}\frac{f\left(\sqrt{\frac{d}{d-2}}G^{-1}_{1-p}\right)}{p}\sqrt{\frac{d-2}{d}},$$

where G^{-1}_{1-p} is the $(1-p)$-th quantile of the distribution of ε. In particular, $G^{-1}_{1-p} = \sqrt{\frac{d-2}{d}}t^{-1}_{d,1-p}$, where $t^{-1}_{d,1-p}$ is the $(1-p)$-th quantile of the distribution of t_d.

In practice, we cannot compute the true values of VaR^p_{T+1} and ES^p_{T+1}, since they depend on the characteristics of the data generating process (i.e. they depend on G and on the conditional variance model σ^2_{T+1}). Thus, we need to estimate these measures, which introduces estimation risk. Our ultimate goal in this paper is to quantify the estimation risk by constructing a confidence – or prediction – interval for the true but unknown risk measures.

We will consider six different estimation methods, divided into three groups.

2.1 Historical Simulation

The first and most commonly used method is referred to as Historical Simulation (HS). It calculates VaR and ES using the empirical distribution of past losses. In particular, the HS estimate of VaR^p_{T+1} is given by

$$HS\text{-}VaR^p_{T+1} = Q_{1-p}\left(\{L_t\}\right),$$

where $Q_{1-p}(\{L_t\})$ denotes the $(1-p)$-th empirical quantile of the losses data $\{L_t\}^T_{t=1}$. In the simulations below we compute the empirical quantiles by linear interpolation between adjacent ordered sample values. The HS estimate of ES^p_{T+1} is given by

$$HS\text{-}ES^p_{T+1} = \frac{1}{\#\left(L_t > HS\text{-}VaR^p_{T+1}\right)} \left(\sum_{L_t > HS\text{-}VaR^p_{T+1}} L_t \right),$$

where $\#\left(L_t > HS\text{-}VaR^p_{T+1}\right)$ denotes the number of observations of $\{L_t\}_{t=1}^T$ that are above the HS estimate of the VaR.

The HS method is completely nonparametric and does not depend on any distributional assumption, thus capturing the nonnormality in the data. It nevertheless ignores the potentially useful information in the volatility dynamics.

The estimation methods that we consider next take into account the volatility dynamics by explicitly relying on the GARCH(1,1) model for predicting σ_{T+1}. In particular, given (4) and (5), estimates of VaR^p_{T+1} and ES^p_{T+1} can be obtained in three steps:

1. Estimate the GARCH(1,1) parameters through Gaussian QMLE, maximizing

$$\ln L \;\propto\; -\frac{1}{2}\sum_{t=1}^{T} \ln\left(\sigma_t^2\right) + \left(\frac{L_t}{\sigma_t}\right)^2 .$$

   Given the QML estimates $\left(\hat{\omega}, \hat{\alpha}, \hat{\beta}\right)$, we can compute the variance sequence $\hat{\sigma}_t^2$ and the implied residuals $\hat{\varepsilon}_t = L_t/\hat{\sigma}_t$ from the past observed squared losses and the past estimated variance using the recursion

$$\hat{\sigma}^2_{t+1} = \hat{\omega} + \hat{\alpha} L_t^2 + \hat{\beta}\hat{\sigma}_t^2,$$

   where $\hat{\sigma}_1^2 = \frac{\hat{\omega}}{1-\hat{\alpha}-\hat{\beta}}$, the unconditional variance of L_t. A prediction of σ_{T+1} is given by $\hat{\sigma}_{T+1}$, where

$$\hat{\sigma}^2_{T+1} = \hat{\omega} + \hat{\alpha} L_T^2 + \hat{\beta}\hat{\sigma}_T^2.$$

2. Choose values for the constants $c_{1,p}$ and $c_{2,p}$. Call these $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$, respectively.

3. Compute the estimates of VaR^p_{T+1} and ES^p_{T+1} as

$$\begin{aligned} \widehat{VaR}^p_{T+1} &= \hat{\sigma}_{T+1}\hat{c}_{1,p} \\ \widehat{ES}^p_{T+1} &= \hat{\sigma}_{T+1}\hat{c}_{2,p}. \end{aligned}$$

We can distinguish between two groups of methods according to rule used to choose the constants $c_{1,p}$ and $c_{2,p}$ in step 2: the normal model and nonparametric methods.

2.2 Normal Conditional Distribution

Erroneously imposing the normal distribution on the innovation term ε_t gives the following estimates of VaR_{T+1}^p and ES_{T+1}^p:

$$\begin{aligned} N\text{-}VaR_{T+1}^p &= \hat{\sigma}_{T+1}\hat{c}_{1,p}^N \\ N\text{-}ES_{T+1}^p &= \hat{\sigma}_{T+1}\hat{c}_{2,p}^N, \end{aligned}$$

where

$$\begin{aligned} \hat{c}_{1,p}^N &= \Phi_{1-p}^{-1}, \\ \hat{c}_{2,p}^N &= \frac{\phi\left(\Phi_{1-p}^{-1}\right)}{p}, \end{aligned}$$

with Φ_{1-p}^{-1} the $(1-p)$-th quantile of a standard normal distribution. We will call this the "Normal" method. This method imposes conditional normality, which does not hold for real data, and it is included only for comparison purposes.

2.3 Nonparametric Methods

These methods estimate $c_{1,p}$ and $c_{2,p}$ using the implied GARCH(1,1) residuals $\hat{\varepsilon}_t = L_t/\hat{\sigma}_t$. They differ in the way they use the residuals to compute $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$.

Extreme Value Theory

The Extreme Value Theory (EVT) approach estimates $c_{1,p}$ and $c_{2,p}$ under the assumption that the tail of the conditional distribution of the GARCH innovation is well approximated by an heavy-tailed distribution. This approach was proposed by McNeil and Frey (2000), who derived estimates of $c_{1,p}$ and $c_{2,p}$ based on the maximum likelihood estimator of the parameters of a Generalized Pareto Distribution (GPD).

Here, we suppose that the tail of the conditional distribution of ε_t is well approximated by the distribution function

$$F(z) = 1 - L(z)\, z^{-1/\xi} \approx 1 - cz^{-1/\xi},$$

whenever $\varepsilon_t > u$, where $L(z)$ is a slowly varying function that we approximate with a constant c, and ξ is a positive parameter. u is a threshold value such that all observations above u will be used in the estimation of ξ. We let T_u denote the number of observations that exceed u. The

Hill estimator (Hill, 1975) $\hat{\xi}$ corresponds to the MLE estimator of ξ under the assumption that the standardized residuals $\hat{\varepsilon}_t$ are approximately i.i.d. It is defined as

$$\hat{\xi} = \frac{1}{T_u} \sum_{t=1}^{T_u} \ln\left(\hat{\varepsilon}_{(T-T_u+t)}\right) - \ln(u),$$

where $\hat{\varepsilon}_{(t)}$ denote the t-th order statistic of $\hat{\varepsilon}_t$ (i.e $\hat{\varepsilon}_{(t)} \geq \hat{\varepsilon}_{(t-1)}$ for $t = 2, \ldots, T$). The important choice of T_u will be discussed at the beginning of the Monte Carlo Results section below.

Given $\hat{\xi}$, an estimate of the tail distribution F is obtained by choosing $c = \frac{T_u}{T} u^{1/\hat{\xi}}$, which derives from imposing the condition $1 - F(u) = \frac{T_u}{T}$. We thus obtain the following estimate of F:

$$\hat{F}(z) = 1 - \frac{T_u}{T}\left(\frac{z}{u}\right)^{-1/\hat{\xi}}.$$

The EVT approach relies on $\hat{F}(z)$ to estimate the constants $c_{1,p}$ and $c_{2,p}$. In particular, the estimate of $c_{1,p}$ is equal to $\hat{F}_{1-p}^{-1}$, the $(1-p)$th-quantile of the tail distribution $\hat{F}$. We can show that

$$\hat{c}_{1,p}^{Hill} = u\left(p\frac{T}{T_u}\right)^{-\hat{\xi}}.$$

Similarly, to compute an estimate of $c_{2,p}$ we use $\hat{F}(z)$ to compute $E\left(\varepsilon | \varepsilon > \hat{F}_{1-p}^{-1}\right)$, where $\varepsilon \sim$ i.i.d. $\hat{F}$. We can show that the following closed form expression holds true

$$E\left(\varepsilon | \varepsilon > \hat{F}_{1-p}^{-1}\right) = \frac{\hat{F}_{1-p}^{-1}}{1-\hat{\xi}}.$$

This implies the following Hill's estimate of $c_{2,p}$:

$$\hat{c}_{2,p}^{Hill} = \frac{\hat{c}_{1,p}^{Hill}}{1-\hat{\xi}}.$$

The Hill's estimates of VaR_{T+1}^{p} and ES_{T+1}^{p} are given by

$$\begin{aligned} Hill\text{-}VaR_{T+1}^{p} &= \hat{\sigma}_{T+1}\hat{c}_{1,p}^{Hill} \\ Hill\text{-}ES_{T+1}^{p} &= \hat{\sigma}_{T+1}\hat{c}_{2,p}^{Hill}, \end{aligned}$$

respectively.

Gram-Charlier and Cornish-Fisher Expansions

This method relies on the Cornish-Fisher and Gram-Charlier expansions to approximate the conditional density of the standardized losses ε_t. For a standardized random variable, a Gram-Charlier expansion produces an approximate density function that can be viewed as an expansion of the standard normal density augmented with terms that capture the effects of skewness and excess kurtosis. Thus, Gram-Charlier expansions are a convenient tool to account for departures of conditional normality.[3]

The Cornish-Fisher expansion approximates the inverse cumulative density function directly. The approximation to $c_{1,p}$ is thus:

$$CF_{1-p}^{-1} = \Phi_{1-p}^{-1} + \frac{\gamma_1}{6}\left[\left(\Phi_{1-p}^{-1}\right)^2 - 1\right] + \frac{\gamma_2}{24}\left[\left(\Phi_{1-p}^{-1}\right)^3 - 3\Phi_{1-p}^{-1}\right] - \frac{\gamma_1^2}{36}\left[2\left(\Phi_{1-p}^{-1}\right)^3 - 5\Phi_{1-p}^{-1}\right],$$

where

$$\begin{aligned}\gamma_1 &= E\left(\varepsilon^3\right)\\ \gamma_2 &= E\left(\varepsilon^4\right) - 3,\end{aligned}$$

with $\varepsilon \sim G(0,1)$. We will refer to the expansions methods generically as GC (for Gram-Charlier). Thus, we have

$$\hat{c}_{1,p}^{GC} = \widehat{CF}_{1-p}^{-1},$$

where $\widehat{CF}_{1-p}^{-1}$ is the sample analogue of CF_{1-p}^{-1}, i.e. it replaces γ_1 and γ_2 with their sample analogues evaluated on the standardized residuals $\hat{\varepsilon}_t = L_t/\hat{\sigma}_t$:

$$\begin{aligned}\hat{\gamma}_1 &= \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^3\\ \hat{\gamma}_2 &= \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^4 - 3.\end{aligned}$$

Thus, we obtain the following estimate of VaR_{T+1}^p:

$$GC\text{-}VaR_{T+1}^p = \hat{\sigma}_{T+1}\hat{c}_{1,p}^{GC}.$$

Similarly, we can define an approximation to $c_{2,p}$ that relies on the Gram-Charlier and Cornish-Fisher expansions. In particular, we can show that

$$c_{2,p}^{GC} = E\left(\varepsilon | \varepsilon > CF_{1-p}^{-1}\right) = \frac{\phi\left(CF_{1-p}^{-1}\right)}{p}\left(1 + \frac{\gamma_1}{6}\left[\left(CF_{1-p}^{-1}\right)^2 - 1\right] + \frac{\gamma_2}{24}CF_{1-p}^{-1}\left[\left(CF_{1-p}^{-1}\right)^2 - 3\right]\right).$$

The Gram-Charlier estimate of ES_{T+1}^p is given by

$$GC\text{-}ES_{T+1}^p = \hat{\sigma}_{T+1}\hat{c}_{2,p}^{GC},$$

[3] For an application of Gram-Charlier expansions in finance, see Backus, Foresi, Li and Wu (1997) and references therein.

where $\hat{c}_{2,p}^{GC}$ is obtained from $c_{2,p}^{GC}$ by replacing CF_{1-p}^{-1}, γ_1 and γ_2 with their sample analogues.

When G is the standard normal distribution, the Gram-Charlier estimates of VaR and ES coincide with those obtained with the "Normal" method.

Filtered Historical Simulation

The Filtered Historical Simulation (FHS) method estimates $c_{1,p}$ and $c_{2,p}$ from the empirical distribution of the (centered) residuals. Thus it combines a model-based variance with a data-based conditional quantile. Several papers including Hull and White (1998), Barone-Adesi et al (1999), and Pritsker (2001) have found the FHS method to perform well.

The FHS estimates of $c_{1,p}$ and $c_{2,p}$ are given by

$$\hat{c}_{1,p}^{FHS} = Q_{1-p}\left(\left\{\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}\right\}_{t=1}^{T}\right)$$

and

$$\hat{c}_{2,p}^{FHS} = \frac{1}{\#\left(\hat{\varepsilon}_t - \overline{\hat{\varepsilon}} > \hat{c}_{1,p}^{FHS}\right)} \left(\sum_{\hat{\varepsilon}_t > \hat{c}_{1,p}^{FHS}} \left(\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}\right)\right),$$

where $\overline{\hat{\varepsilon}} = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t$. Centered residuals are considered because their sample average is zero by construction, thus better mimicking the true mean zero expectation of the standardized errors ε_t. If a constant is included in the losses model, $\sum_{t=1}^{T}\hat{\varepsilon}_t = 0$ and centering of the residuals becomes irrelevant.

This implies the following FHS estimates of VaR_{T+1}^p and ES_{T+1}^p:

$$\begin{aligned} FHS\text{-}VaR_{T+1|T}^p &= \hat{\sigma}_{T+1}\hat{c}_{1,p}^{FHS} \\ ES\text{-}VaR_{T+1}^p &= \hat{\sigma}_{T+1}\hat{c}_{2,p}^{FHS}. \end{aligned}$$

3 Resampling Methods for Estimation Risk

In this section we describe the bootstrap methods we use to assess the estimation risk in the risk estimates presented above.

Our first bootstrap method applies to Historical Simulation. This bootstrap method ignores any volatility dynamics and simply treats losses as being i.i.d. This "naive" bootstrap method generates pseudo losses by resampling with replacement from the set of original losses, according to the following algorithm:

Bootstrap Algorithm for Historical Simulation Risk Measures

Step 1. Generate a sample of T bootstrapped losses $\{L_t^* : t = 1, \ldots, T\}$ by resampling with replacement from the original data set $\{L_t\}$.

Step 2. Compute the HS estimates of VaR and ES on the bootstrap sample:

$$HS\text{-}VaR_{T+1}^{*p} = Q_p\left(\{L_t^*\}_{t=1}^T\right).$$

$$HS\text{-}ES_{T+1}^{*p} = \frac{1}{\#\left(L_t^* > HS\text{-}VaR_{T+1}^{*p}\right)}\left(\sum_{L_t^* > HS\text{-}VaR_{T+1}^{*p}} L_t^*\right).$$

Step 3. Repeat Steps 1 and 2 a large number of times, B say, and obtain a sequence of bootstrap HS risk measures. For instance, $\left\{HS\text{-}VaR_{T+1}^{*p(i)} : i = 1, \ldots, B\right\}$ denotes a sequence of bootstrap VaR measures. We set $B = 999$ in our Monte Carlo simulations below.

Step 4. The $100\,(1-\alpha)\,\%$ bootstrap prediction interval for VaR_{T+1}^p is given by

$$\left[Q_{\alpha/2}\left(\left\{HS\text{-}VaR_{T+1}^{*p(i)}\right\}_{i=1}^B\right), Q_{1-\alpha/2}\left(\left\{HS\text{-}VaR_{T+1}^{*p(i)}\right\}_{i=1}^B\right)\right],$$

where $Q_\alpha(\cdot)$ is the $\alpha-$quantile of the empirical distribution of $\left\{HS\text{-}VaR_{T+1}^{*p(i)}\right\}$. A similar bootstrap interval can be computed for ES_{T+1}^p.

Following the Historical Simulation approach, this naive bootstrap method is completely non-parametric, avoiding any distributional assumptions on the data. However, by implicitly assuming that returns are i.i.d., this method fails to capture the dependence in returns when it exists. In particular, as our simulations will show, this method of computing confidence intervals for risk measures is not appropriate when returns follow a GARCH model.

The validity of the bootstrap for financial data depends crucially on its ability to correctly mimic the dependence properties of returns. A natural and often used bootstrap method for GARCH models consists of resampling with replacement the standardized residuals, the idea being that the standardized errors are i.i.d. in the population. The bootstrap returns are then recursively generated using the GARCH volatility dynamic equation and the resampled standardized residuals. The bootstrap methods that we describe next are based on this general idea.

As described in the previous section, under model (1), the VaR and ES have the following simplified expressions

$$VaR_{T+1}^p = \sigma_{T+1} c_{1,p}, \tag{6}$$

and

$$ES^p_{T+1} = \sigma_{T+1} c_{2,p}, \tag{7}$$

where $c_{1,p}$ and $c_{2,p}$ are a function of G and p, and σ_{T+1} is given by the square root of

$$\sigma^2_{T+1} = \omega + \alpha L^2_T + \beta \sigma^2_T. \tag{8}$$

Given (6) and (7), there are two sources[4] of risk associated with predicting VaR^p_{T+1} and ES^p_{T+1} using information available at T. One is the uncertainty in computing $c_{1,p}$ and $c_{2,p}$. If the risk model correctly specifies G, then this source of risk is not present. The other source of risk relates to predicting the volatility σ_{T+1} using day T's information. For our GARCH(1,1) model, it is easy to see that σ^2_{T+1} depends on information available at day T and on the unknown parameters ω, α and β. In particular, using the GARCH equation (8), we can write σ^2_T as a function of past losses as follows:

$$\sigma^2_T = \frac{\omega}{1-\alpha-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j \left(L^2_{T-j-1} - \frac{\omega}{1-\alpha-\beta} \right).$$

Replacing ω, α and β with their MLE estimates yields

$$\hat{\sigma}^2_T = \frac{\hat{\omega}}{1-\hat{\alpha}-\hat{\beta}} + \hat{\alpha} \sum_{j=0}^{T-2} \hat{\beta}^j \left(L^2_{T-j-1} - \frac{\hat{\omega}}{1-\hat{\alpha}-\hat{\beta}} \right), \tag{9}$$

which delivers a point estimate $\hat{\sigma}^2_{T+1} = \hat{\omega} + \hat{\alpha} L^2_T + \hat{\beta} \hat{\sigma}^2_T$. The need to estimate the GARCH parameters introduces the second source of estimation risk.

The presence of estimation risk in computing VaR^p_{T+1} and ES^p_{T+1} is our main motivation for using the bootstrap to obtain prediction intervals for these risk measures. The bootstrap methods we use are based on Pascual, Romo and Ruiz (2001), who proposed a bootstrap method for building prediction intervals for returns volatility σ_t based on the GARCH(1,1) model. In particular, for the nonparametric methods, we extend the Pascual, Romo and Ruiz (2001) resampling scheme to the case of VaR^p_{T+1} and ES^p_{T+1} by using the bootstrap to account for estimation error not only in σ_{T+1} but also in the constants $c_{1,p}$ and $c_{2,p}$ that multiply σ_{T+1}.

Bootstrap Algorithm for GARCH-Based Measures of Risk

Step 1. Estimate the GARCH model by MLE and compute the centered residuals $\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}$, where $\hat{\varepsilon}_t = \frac{L_t}{\hat{\sigma}_t}$, $t = 1, \ldots, T$. Let $\hat{G}_T$ denote the empirical distribution function of $\hat{\varepsilon}_t$.

[4] In general, model risk is a third source of uncertainty when forecasting VaR^p_{T+1} and ES^p_{T+1}. Here, we abstract from this source of uncertainty since we take the GARCH model of returns as being correctly specified.

Step 2. Generate a bootstrap pseudo series of portfolio losses $\{L_t^* : t = 1, \ldots, T\}$ using the recursions

$$\begin{aligned} \hat{\sigma}_t^{*2} &= \hat{\omega} + \hat{\alpha} L_{t-1}^{*2} + \hat{\beta} \hat{\sigma}_{t-1}^{*2}, \\ L_t^* &= \hat{\sigma}_t^* \varepsilon_t^*, \text{ for } t = 1, \ldots, T \end{aligned}$$

where $\varepsilon_t^* \sim$ i.i.d.$\hat{G}_T$ and where $\hat{\sigma}_1^{*2} = \hat{\sigma}_1^2 = \frac{\hat{\omega}}{1-\hat{\alpha}-\hat{\beta}}$. With the bootstrap pseudo-data $\{L_t^*\}$, compute the bootstrap MLE's $\hat{\omega}^*, \hat{\alpha}^*$ and $\hat{\beta}^*$.

Step 3. Obtain a bootstrap prediction of volatility $\hat{\sigma}_{T+1}^*$ according to

$$\hat{\sigma}_{T+1}^{*2} = \hat{\omega}^* + \hat{\alpha}^* L_T^{*2} + \hat{\beta}^* \hat{\sigma}_T^{*2},$$

given the initial values

$$\begin{aligned} L_T^* &= L_T, \\ \hat{\sigma}_T^{*2} &= \frac{\hat{\omega}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*} + \hat{\alpha}^* \sum_{j=0}^{T-2} \hat{\beta}^{*j} \left(L_{T-j-1}^2 - \frac{\hat{\omega}^*}{1 - \hat{\alpha}^* - \hat{\beta}^*} \right). \end{aligned} \quad (10)$$

Step 4. Compute $\hat{c}_{1,p}^*$ and $\hat{c}_{2,p}^*$, the bootstrap estimates of $c_{1,p}$ and $c_{2,p}$. These bootstrap estimates are computed exactly in the same fashion as $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$ with the difference that they are evaluated on the bootstrap data instead of the real data. In particular, for the Normal model we simply set

$$\hat{c}_{1,p}^* = \hat{c}_{1,p}^N \text{ and } \hat{c}_{2,p}^* = \hat{c}_{2,p}^N$$

where $\hat{c}_{1,p}^i$ and $\hat{c}_{2,p}^i$ are as described before. In contrast, for the nonparametric methods, we first compute the bootstrap residuals

$$\hat{\varepsilon}_t^* = \frac{L_t^*}{\hat{\sigma}_t^*},$$

with $\hat{\sigma}_t^{*2} = \hat{\omega}^* + \hat{\alpha}^* R_{t-1}^{*2} + \hat{\beta}^* \hat{\sigma}_{t-1}^{*2}$ and $\hat{\sigma}_1^{*2} = \hat{\sigma}_1^2$. Next, we evaluate the estimates of $c_{1,p}$ and $c_{2,p}$ on the data set $\{\hat{\varepsilon}_t^*\}_{t=1}^T$. For instance,

$$\hat{c}_{1,p}^{*FHS} = Q_{1-p}\left(\left\{ \hat{\varepsilon}_t^* - \overline{\hat{\varepsilon}^*} \right\}_{t=1}^T \right).$$

Step 5. For each estimation method, compute the bootstrap estimates of VaR_{T+1}^p and ES_{T+1}^p using $\hat{\sigma}_{T+1}^*$ and $\hat{c}_{1,p}^*$ and $\hat{c}_{2,p}^*$.

Step 6. Identical to steps 3 and 4 in the naive bootstrap.

Step 3 accounts for the estimation risk in computing $\hat{\sigma}_{T+1}$ by replacing the estimates $\hat{\omega}, \hat{\alpha}$ and $\hat{\beta}$ by their bootstrap analogues $\hat{\omega}^*, \hat{\alpha}^*$ and $\hat{\beta}^*$ when computing $\hat{\sigma}^*_{T+1}$. In particular, (10) replicates the way in which $\hat{\sigma}^2_T$ is computed in (9). Notice however that $\hat{\sigma}^{*2}_T$ is conditional on the observed past observations on the losses $\{L_t : t = 1, \ldots, T\}$, not on the bootstrap losses generated in step 2, implying that it is small when the (true) losses are small at the end of the sample and large when they are large.

For the FHS method, bootstrap residuals are centered before computing the empirical quantile as a way to enforce the mean zero property on the estimated bootstrap residuals (centering of the residuals is not needed if a constant is included in the returns model since in that case the residuals have mean zero by construction).

We conclude this section by noting that it may be possible to apply asymptotic approximations such as the delta-method to calculate prediction intervals for the GARCH variance forecast.[5] However, it is not at all obvious how to calculate prediction intervals for VaR and ES using the delta method in the nonparametric risk models we consider. Furthermore, even in parametric cases, the approximate delta-method is likely to perform worse than the resampling techniques considered here. In the following we therefore restrict attention to prediction intervals calculated via our resampling technique.

4 Monte Carlo Results

As indicated in the introduction the purpose of our paper is twofold: First, we want to assess the potential loss of accuracy from estimation error when calculating VaR and ES. Second, we want to assess our ability to quantify ex-ante the magnitude of this error via the construction of confidence intervals around the risk measures. This section provides quantitative evidence on these two issues through a Monte Carlo study. The main focus of our analysis will be the realistic situation of time varying portfolio risk driven in our case by a GARCH model. However, before venturing into the more complicated GARCH case it is sensible to apply our analysis to the case of simple, independent losses.

4.1 Independent Losses

In Table 1 we simulate independent daily loss data from a Student distribution with mean zero and variance $20^2/252$, implying a volatility of 20% per year. and calculate VaR (top panel) and ES

[5] This approach is taken for example in Duan (1994).

(bottom panel) risk measures by Historical Simulation.[6] Each line in the table corresponds to one of four experiments with degrees of freedom equal to 8 or 500, and estimation sample sizes equal to 500 or 1000 days respectively. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel).

The top left panel shows that the HS-VaRs have little bias but the root mean squared errors (RMSEs) indicate that the VaRs are somewhat imprecisely estimated. The RMSE is around 10% of the true VaR value when the degree of freedom equals 8. The top right panel show that the VaR confidence intervals from the bootstrap have nominal coverage rates close to the promised 90%. The average width of the bootstrap intervals is between 17 and 37 percent of the true VaR value depending on the sample size and on the degrees of freedom. In the most realistic case where $d = 8$ and $T = 500$ the average 90% interval width is a substantial 37% of the true VaR value.

The bottom left panel shows that the bias of the ES point estimates are small but again that the RMSEs are substantial in the leading case where $d = 8$ and $T = 500$. Furthermore the bottom right panel shows that the coverage rates of the 90% confidence intervals are substantially less than 90%. The confidence intervals can therefore not be trusted for the ES risk measures. This finding is repeated often in the GARCH analysis below.

4.2 GARCH Losses

We will now consider four versions of the GARCH-$t(d)$ data generating process (DGP) below. In each version we set $\alpha = .10$ and $\omega = \left(20^2/252\right) * (1 - \alpha - \beta)$. The unconditional volatility is thus 20% per year. Our four chosen parameterizations are:

1) Benchmark: $\beta = .80, d = 8$

2) High Persistence: $\beta = .89, d = 8$

3) Low Persistence: $\beta = .40, d = 8$

4) Normal Distribution: $\beta = .80, d = 500$

Recall that before applying the Hill estimator for the extreme value distribution we need to choose a cut-off point, T_u, which defines the sub-sample of extremes from which the tail index parameter will be estimated. In order to pick this important parameter we perform an initial Monte Carlo experiment in which we simulate data from the four DGPs above, estimate the tail index on a grid of cut-off values, and finally compute the resulting bias and root mean squared error measures (RMSEs) from the one-day VaR and ES forecasts. Figures 1 and 2 show the results for the case of 500 and 1000 total estimation sample points respectively. In each case, we choose

[6] We only analyze the Historical Simulation risk model here as the GARCH based risk models are not identified when returns are independent.

a grid of truncation points which correspond to including the 0.5% to 10% largest losses in the sub-sample of extremes. The horizontal axis in each figure denotes the number of included extreme observations (out of 500 and 1000 respectively), and the vertical axis shows the bias and RMSEs. From the viewpoint of minimizing RMSE subject to achieving a bias that is close to zero, and looking broadly across the four DGPs, it appears that a percentage cut-off of 2% is reasonable for both VaR and ES. Notice that we do not want to choose the truncation point on a case by case basis as that would potentially bias the overall results in favor of the Hill-based risk model.

Tables 2-5 contain the Monte Carlo results corresponding to the four DGPs above. The top half of each table contains the VaR results and the bottom half the ES results. The left half of each table contains the accuracy properties of the point estimates of the relevant risk measure and the right half contains the 90% bootstrap interval properties. For both the VaR and ES forecasts we consider two estimation sample sizes, $T = \{500, 1000\}$.

In all the experiments we calculate the properties of the point estimates from 100,000 Monte Carlo replications. For the properties of the bootstrap prediction intervals, we consider only 5,000 Monte Carlo replications, each with 999 bootstrap replications. Any Monte Carlo study of the bootstrap is computationally demanding and this is particularly so in our study due to the nonlinear optimization involved in estimating GARCH.

4.3 Point Predictions of VaR and ES

While the main focus of our paper is on constructing finite sample prediction intervals of the VaR and ES measures, we first consider the various models' ability to accurately point forecast the risk measures. The point prediction results on VaR and ES are reported in terms of bias and root mean squared error, which are reported in the left half of each table.

4.3.1 The Benchmark Case

The top panel of Table 2 contains the VaR results for our benchmark DGP when the sample size is $T = 500$. Considering first the bias of the VaR estimates, the main thing to note is the upward bias of the HS and the downward bias of the Normal. The latter is of course to be expected as the Normal imposes a distribution tail which is too thin for the 1% coverage rate. The other models appear to show only minor biases with the FHS model displaying the smallest bias overall.

In terms of the root mean squared error (RMSE) of the VaR estimates, we see that the HS has by far the highest RMSE, followed by the GC model. The Hill model in particular, but also the FHS model are much lower. The RMSE of the Normal is also low but, as mentioned before, it displays considerable bias.

Increasing the sample size to 1000 in the second panel of Table 2 implies smaller biases in general. The HS is still biased upwards and the Normal downwards. In terms of RMSE, the Hill and FHS methods perform very well.

We next examine the quality of the point predictions of ES by the various models. We now find a very large downward bias for the GC and again for the Normal model. In comparison with the VaR results, the various estimated ES models have RMSEs which are considerably larger. The increase in RMSE is due partly to increases in the bias. The results for the GC model indicate that it is not useful for ES calculations the way we have implemented it here. Notice that in the ES case the GC model is an aggregate of two approximations: First, the Cornish-Fisher approximation to the VaR and second the Gram-Charlier approximation to the cumulative density. Unfortunately, the two approximation errors appear to compound each other for the purpose of ES calculation.

4.3.2 The High Persistence Case

The top half of Table 3 reports the VaR findings for a DGP of high volatility persistence and therefore also high kurtosis. We see that the biases and RMSEs are comparable to the benchmark DGP in Table 2 for the conditional models but not for the HS model. The HS model is now even more biased and has a RMSE of more than 50% of the average true VaR, which is approximately 2.71. The Hill and FHS models again perform very well.

The bottom half of Table 3 reports results for ES using the high volatility persistence DGP. We find that the results are very close to the bottom half of Table 2 for the conditional models but not for HS. This finding matches the results for VaR reported in the top halves of Tables 2 and 3 respectively. As before, the bias and RMSEs of the HS model are very large, and for the ES the GC model again performs poorly.

4.3.3 The Low Persistence Case

In the top half of Table 4 we consider the VaR case of low volatility persistence. Not surprisingly the HS model performs much better now. Interestingly, the Hill and FHS models perform very well here also. The bottom half of the table shows the results for ES forecasting in the low persistence process. As in the VaR case, we see that the HS model now performs relatively well.

4.3.4 The Conditional Normal Case

The top half of Table 5 contains VaR results for the conditionally normal GARCH DGP. Comparing with Table 2 we see that the bias and RMSEs are considerably smaller now. It is still the case that the HS model is much worse than the conditional models. The Normal model of course performs very well now as it is the true model. Interestingly, the Hill and FHS models which do not directly nest the Normal model still perform decently. This is important as the risk manager of course never knows exactly the degree of conditional non-normality in the return distribution.

The bottom half of Table 5 considers the ES risk measure. Comparing the bottom of Table 5 with the bottom of Table 2 we see that the biases and RMSEs are generally much smaller under conditional normality. The biases and RMSEs for ES are very much in line with the ones from VaR in the top half of Table 5. This is sensible from the perspective that under conditional normality the ES does not contribute information over and beyond the VaR.

4.4 Bootstrap Prediction Intervals for VaR and ES

The above discussion was concerned with the precision of the VaR and ES point forecasts. We now turn our attention to the results for the bootstrap prediction intervals from the different VaR and ES models. That is, we want to assess the ability of the bootstrap to reliably predict ex ante the accuracy of each method in predicting the 1-day-ahead 1% VaR and ES. The prediction interval results are reported in the right hand side of each table. We show the true coverage rate of nominal 90% intervals as well as the average limits of the confidence intervals and the average width of the confidence interval as a percentage of the true VaR point forecast.

4.4.1 The Benchmark Case

Turning back to Table 2 and looking at the top panel, we remark that the historical simulation VaR (HS) intervals (calculated from the i.i.d. bootstrap) have a terribly low effective coverage for a promised nominal coverage of 90%. Furthermore, the confidence intervals are on average very wide. The HS method ignores the dynamics in the DGP which is costly both in terms of coverage and width.[7]

The VaR imposing the conditional normal distribution (Normal) has a coverage which is as bad as the HS model but which has a much smaller average width. The small width does of course not offer much comfort here as the nominal coverage is much too small. The GC model has larger coverage than the Hill but has wider intervals. Finally, the FHS model has slight over-coverage,

[7] We also calculated GARCH-bootstrap confidence intervals for the HS model. These performed better than the iid bootstrap intervals reported in the tables but they were still very inaccurate and were therefore not included in the tables. The iid boostrap is shown here because it is arguably most in line with the model-free spirit of the HS model.

which is arguably to be preferred to under-coverage, but it also has a fairly wide average coverage intervals.

In the second panel of Table 2 we increase the risk manager's sample size to 1000 past return observations in each simulation. Comparing with the top panel in Table 2 the result are as follows: The HS model coverage actually gets worse with sample size. In the short (500 observations) sample the HS model is able to pick up some of the dynamics in the return process, but it is less able to do so as the sample size increases. The average width is smaller as the sample size increases due to the higher precision in estimating the (unconditional) VaR. The Normal model also has worse coverage and better width. This may seem puzzling, but note that there is no reason to believe that a larger sample size will improve the coverage of a misspecified model. The Hill and GC models both have better coverages and widths now. Finally, notice that the FHS model also benefits from the larger estimation samples and show better coverages and lower widths.

The bottom half of Table 2 reports results for the bootstrap prediction intervals from the different ES models. We notice the following: The Historical Simulation ES intervals (calculated from the i.i.d. bootstrap) have a low effective coverage for a promised nominal coverage of 90%. Furthermore the confidence intervals on average are quite wide. The HS results for ES are roughly comparable with those for VaR in Table 2. The ES imposing the conditional normal distribution (Normal) has a surprisingly low coverage. Thus, while the normal distribution is bad for VaR prediction intervals it is much worse for ES prediction intervals. The Hill model has the best coverage but is quite wide. The GC model has very low coverage and quite wide intervals. Finally, the FHS model has considerable under-coverage. This is in contrast with the VaR intervals in the top half of the table.

Looking more broadly at the results in Table 2, we see that the Hill model has the best coverage followed by the FHS model. The HS, Normal and GC models have poor coverage. Compared with the top half of the table it thus appears that while the FHS performs well for VaR prediction interval calculation, it is less useful for ES prediction intervals. The Hill estimator appears to be preferable here. Generally the coverage rates are considerably worse for ES than for VaR.

4.4.2 The High Persistence Case

The top right hand side of Table 3 reports VaR interval results from a return generating process with relatively high persistence. Comparing panel for panel with the benchmark process in Table 2, we notice that the HS model has worse coverage and worse width, whereas the Normal model has better coverage. The GC model has similar coverage but wider intervals. The FHS has good coverage under high persistence but the intervals are wider here as well. Thus, the higher persistence associated with higher kurtosis leads to wider prediction intervals overall.

The bottom right hand side of Table 3 reports ES results. Comparing the VaR and ES results in Table 3 we see that the coverage rates are typically much worse for ES than VaR.

A comparison of the results for ES against the benchmark process in Table 2 reveals that the HS model has worse coverage and worse width. The Normal model has very poor coverage still. The Hill model generally has better coverage but wider intervals. The GC model still has very poor coverage. Finally, the FHS has roughly the same coverage under high persistence but the widths are worse here as well. The higher persistence again leads to wider prediction intervals overall.

4.4.3 The Low Persistence Case

The top right hand side of Table 4 reports VaR results from returns with low variance persistence. Not surprisingly the results are reversed from Table 3, which contained high persistence variances. We now find that the HS model has much better coverage and slightly better widths. The low persistence process is closer to i.i.d., the only assumption under which the HS model is truly justified. The Normal model has worse coverages but it has better widths. The Hill and GC models have similar coverages and better widths than before. Finally, the FHS model has worse coverages, but the widths are slightly better.

The bottom right hand side of Table 4 reports ES results from returns with low variance persistence. We now find that the HS performs much better as we are closer to the i.i.d. case but otherwise the results are similar to the benchmarks in Table 2.

4.4.4 The Conditional Normal Case

In Table 5 we generate returns which are close to conditionally normally distributed. Comparing the VaR panels in Table 5 with the corresponding panels in Table 2, where the conditional returns were $t(8)$, we see the following: The HS model now has worse coverage but also lower width than before. The Normal model has better coverage and better width. This is not surprising as the Normal model is now closer to the truth. The Hill and GC models have similar coverage and better width than before. The FHS model also has roughly the same coverage under conditional normality but better width than under the conditional $t(8)$. Not surprisingly, the models generally perform better under conditional normality. It is perhaps surprising that the Hill model performs well under conditional normality as the tail index parameter may be biased in this case.

In the bottom half of Table 5 we report the ES results. As expected, the models generally perform better under conditional normality in terms of coverage. The HS model is again notably worse than the other models, the FHS is also worse than the others. The Normal model and the GC model which nests the normal models naturally have very good coverages.

4.5 Summary of Results

Based on the results in Tables 2-5, we reach the conclusion that the HS model not only gives bad point estimates of VaR and ES estimates (see also Pritsker 2001) but it also implies very poor confidence intervals. This is true even when the degree of volatility persistence is relatively modest. The Normal model of course works reasonably well when the normality assumption is close to true in the data but otherwise not. The Hill and FHS models perform quite well, even for the conditionally normal distribution. We noticed also that the GC model has serious problems when calculating ES point estimates and intervals for conditionally non-normal returns. Finally, the FHS model works particularly well for VaR calculations.

In general we found that the RMSEs were much higher (relative to the true value) when calculating ES compared to VaR measures. Thus, while the ES measure in theory conveys more information about the loss distribution tail, it is also harder to estimate precisely. This point is important to consider when arguing over the relative merits of the two risk measures.

Unfortunately, it is also much harder to reliably assess ex ante the accuracy of ES measures compared with the VaR measures. While the Hill, GC and particularly the FHS model give quite reliable coverage rates for the 90% confidence intervals around the VaR point forecast, the corresponding coverage rates for the ES measure are typically much lower than 90% and thus unreliable. We suspect that the higher bias of the ES forecasts is the culprit of the under-coverage in this case. Notice that from a conservative risk management perspective over-coverage would be preferred to under-coverage.

Finally, while the FHS model appears to be preferable for calculating VaR forecasts and forecasts intervals, the Hill model performs well in the ES case. The distribution-free FHS model is useful for quantile forecasting but when the mean beyond the quantile must be forecast, then the functional form estimation implicit in the Hill method adds value.

5 Conclusions

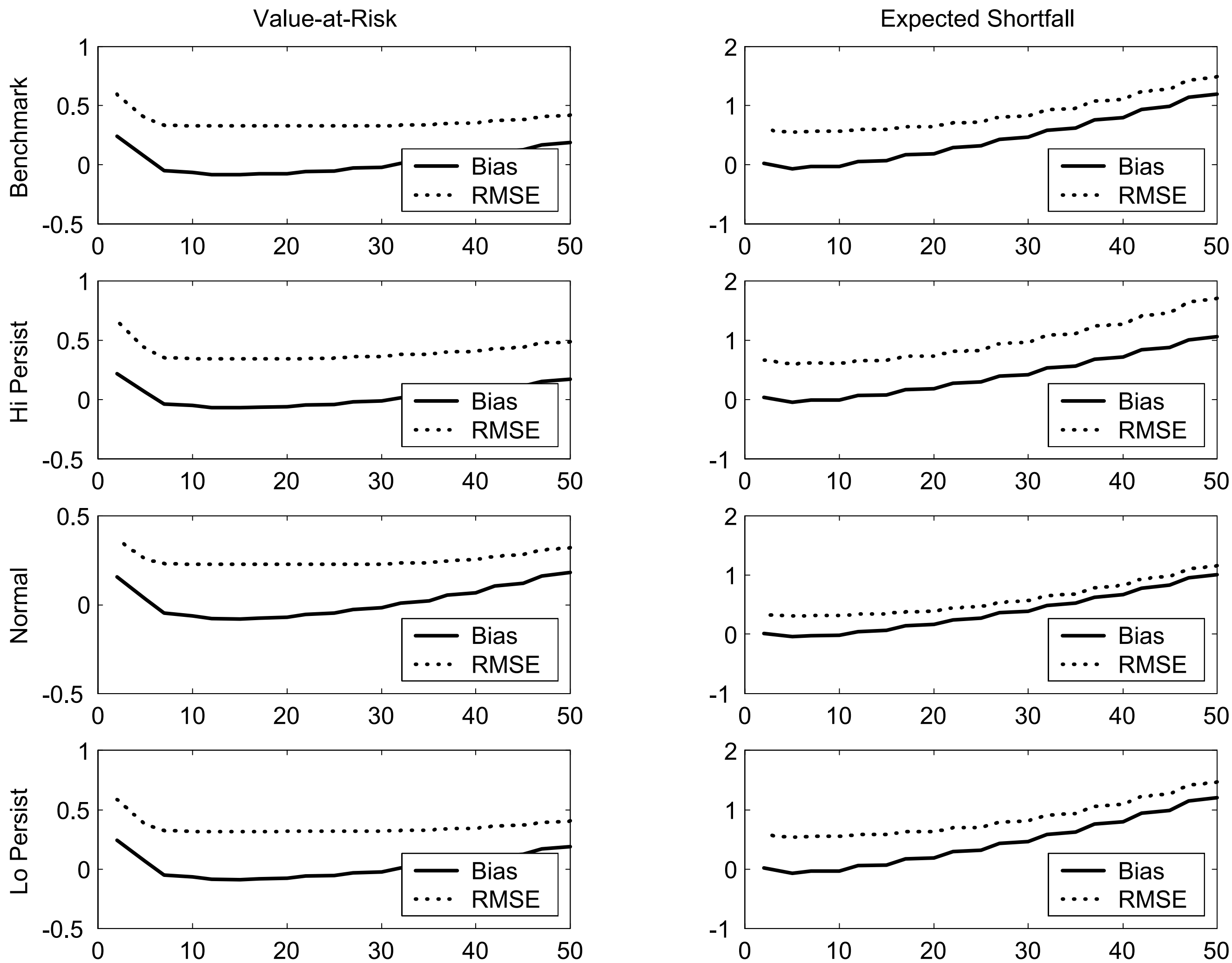
Risk managers and portfolio managers often haggle over the precision of a VaR estimate. A trader faced with a point estimate VaR which exceeds the agreed upon VaR limit may be forced to rebalance the portfolio at an inopportune time. Quantifying the uncertainty of the VaR point estimate is important because it allows for risk managers to make more informed decisions when dictating a portfolio rebalance.

Consequently we suggest a bootstrap method for calculating confidence intervals around the VaR point estimate. The procedure is valid even under conditional heteroskedasticity and nonnormality, which are important features of speculative asset returns. We find that the FHS VaR models yield confidence intervals which have correct coverage but which are also quite wide. In our benchmark case, the average width of the VaR interval for the FHS model is 27-38% of the true VaR depending on the sample size. VaR models based on the normal distribution are much narrower but also often too narrow causing under-coverage of the intervals. We also find that the accuracy of ES forecasts is typically much lower than that of VaR forecasts. Furthermore the accuracy of the ES forecasts is harder to quantify ex ante. In our benchmark case the average width of the ES confidence interval is 22-42% of the true ES value (again depending on the sample size) for the best model. We believe that this quantification of the level of estimation risk in common risk models have important implications for the choice of risk model and risk measure.[8]

We have studied the effects of estimation risk at the portfolio level only (See Benson and Zangari, 1997, Engle and Manganelli, 2004, and Zangari, 1997). Many banks rely instead on multivariate risk factor models such as those considered in Glasserman, Heidelberger, and Shahabuddin (2000 and 2002). The issue of estimation risk is of course equally important but even more complicated in the case of multiple risk factors. We leave this issue for future work.

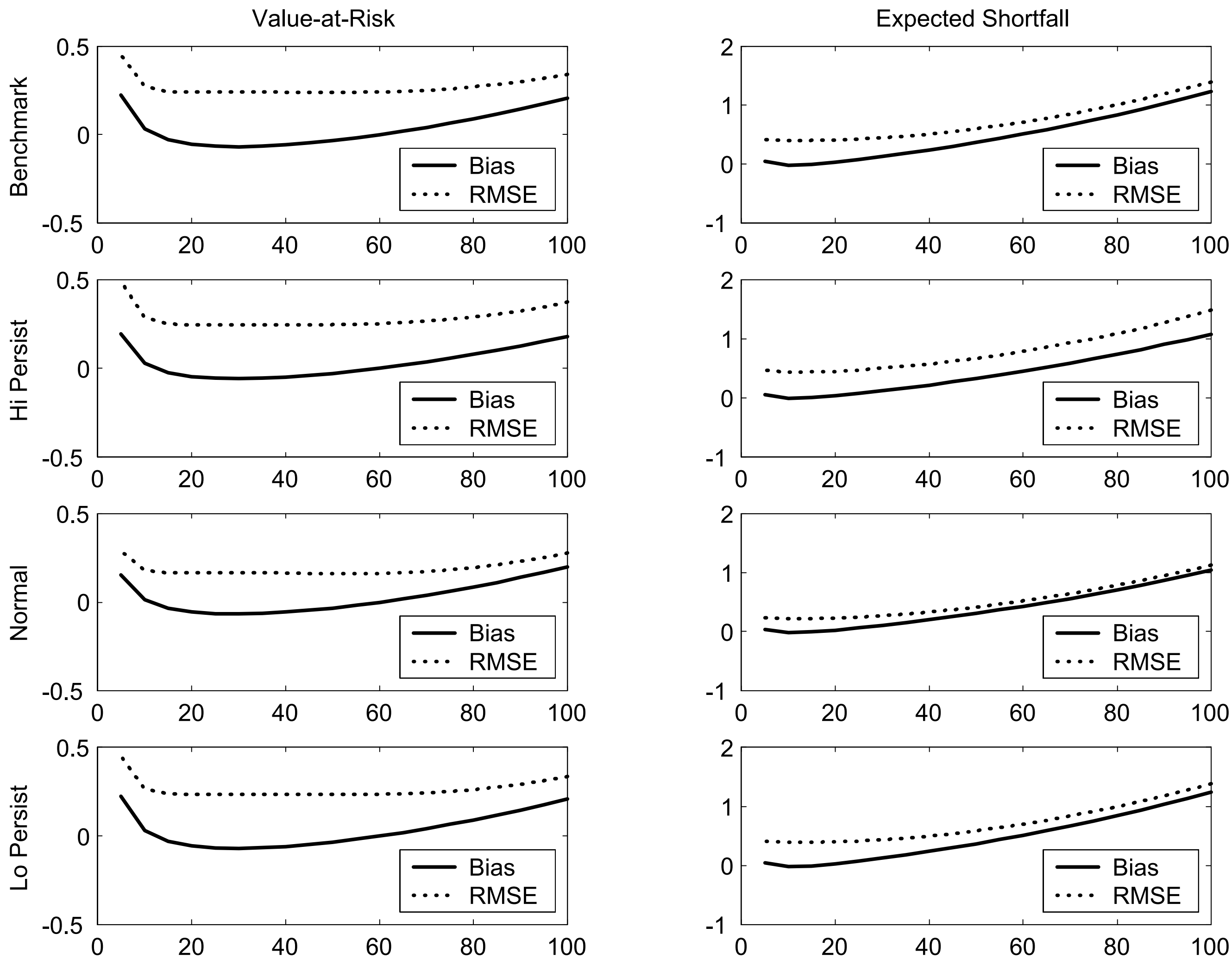
[8] Note that one of the industry benchmarks, namely RiskMetrics, relies on calibrated rather than estimated parameters and does not allow for the calculation of estimation risk. The issue of VaR uncertainty is nevertheless crucial in those models as well but it is not easily quantified.

Figure 1: RMSE and Bias of Hill Estimator for Various Samples of Extremes
The Total Sample Consists of 500 Daily Loss Observations



Notes to Figure: We perform a Monte Carlo study of the choice of sample size of extremes in EVT parameter estimation. The figure shows the root mean squared error (dashed) and bias (solid) of the VaR (left panel) and ES (right panel) estimates against the extremes estimation sample size. The total sample size is 500 observations.

Figure 2: RMSE and Bias of Hill Estimator for Various Samples of Extremes
The Total Sample Consists of 1000 Daily Loss Observations



Notes to Figure: We perform a Monte Carlo study of the choice of sample size of extremes in EVT parameter estimation. The figure shows the root mean squared error (dashed) and bias (solid) of the VaR (left panel) and ES (right panel) estimates against the extremes estimation sample size. The total sample size is 1000 observations.

Table 1: 90% Prediction Intervals for 1% VaR and ES

Historical Simulation Method when Losses are i.i.d.

DGP: $L_t \sim$ i.i.d. $t(d)$ with $E(L_t) = 0$ and $Var(L_t) = \frac{(20)^2}{252}$								
		VaR Properties			VaR Bootstrap Intervals Properties			
d	T	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
8	500	3.200	0.040	0.339	89.44	2.70	3.87	37.18
	1000	3.165	0.004	0.229	88.58	2.81	3.56	23.75
500	500	2.950	0.016	0.221	88.52	2.60	3.35	25.27
	1000	2.929	−0.006	0.152	88.46	2.69	3.18	16.86
		ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
8	500	3.823	−0.095	0.496	71.74	3.08	4.42	34.06
	1000	3.859	−0.058	0.353	79.44	3.30	4.35	26.73
500	500	3.304	−0.053	0.267	75.74	2.86	3.61	22.39
	1000	3.324	−0.033	0.190	81.38	3.00	3.58	17.08

Notes to Table: We simulate T independent daily Student's $t(d)$ losses and calculate VaR (top panel) and ES (bottom panel) risk measures by Historical Simulation. The four experiments correspond to degrees of freedom equal to 8 and 500, and estimation sample sizes equal to 500 and 1000 days. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel). The true VaR values are 3.160 (for $d = 8$) and 2.934 (for $d = 500$). The true ES values are 3.918 (for $d = 8$) and 3.357 (for $d = 500$).

Table 2. 90% Prediction Intervals for 1% VaR and ES: Benchmark GARCH Case

DGP: GARCH-$t\,(d)$ with $\alpha = 0.10, \beta = 0.80$ and $d = 8$

T	Method	VaR Properties			VaR Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.282	0.175	0.748	61.00	2.73	4.02	41.65
	Normal	2.866	−0.240	0.331	60.18	2.53	3.18	20.99
	Hill	3.043	−0.064	0.327	84.88	2.55	3.50	30.79
	GC	3.194	0.088	0.493	85.20	2.60	3.66	34.13
	FHS	3.138	0.032	0.383	91.32	2.57	3.76	38.40
1000	HS	3.240	0.134	0.671	47.64	2.85	3.69	27.07
	Normal	2.872	−0.234	0.289	41.22	2.63	3.09	14.86
	Hill	3.051	−0.055	0.238	84.94	2.69	3.38	22.09
	GC	3.245	0.139	0.435	87.46	2.77	3.62	27.39
	FHS	3.106	0.000	0.268	90.58	2.70	3.52	26.65

T	Method	ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.966	0.115	0.978	60.86	3.15	4.60	37.76
	Normal	3.283	−0.568	0.631	19.10	2.90	3.64	19.39
	Hill	3.806	−0.046	0.561	81.60	2.99	4.60	41.73
	GC	2.609	−1.242	1.435	41.58	1.98	3.79	47.02
	FHS	3.728	−0.123	0.539	74.62	2.95	4.35	36.50
1000	HS	4.020	0.169	0.893	53.34	3.40	4.56	30.28
	Normal	3.289	−0.561	0.601	6.22	3.01	3.54	13.73
	Hill	3.865	0.014	0.411	87.18	2.69	3.38	22.09
	GC	2.490	−1.360	1.484	12.94	1.99	3.33	34.64
	FHS	3.771	−0.079	0.394	79.30	3.18	4.27	28.23

Notes to Table: We simulate T daily GARCH(1,1) losses with Student's $t(d)$ innovations (benchmark parameter configuration) and calculate VaR (top panel) and ES (bottom panel) risk measures by various methods. The two experiments correspond to estimation sample sizes equal to 500 and 1000 days. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel).

Table 3. 90% Prediction Intervals 1% VaR and ES: High Persistence

DGP: GARCH-$t\,(d)$ with $\alpha = 0.10, \beta = 0.89$ and $d = 8$

		VaR Properties			VaR Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.242	0.537	1.958	32.98	2.61	4.04	53.03
	Normal	2.471	−0.235	0.341	61.30	2.11	2.81	25.91
	Hill	2.654	−0.051	0.345	84.56	2.14	3.10	35.42
	GC	2.771	0.066	0.556	86.34	2.18	3.30	41.63
	FHS	2.738	0.033	0.418	90.82	2.17	3.32	43.03
1000	HS	3.330	0.619	2.064	20.68	2.84	3.91	40.34
	Normal	2.485	−0.226	0.299	45.86	2.22	2.65	16.35
	Hill	2.663	−0.048	0.247	85.14	2.27	2.89	23.41
	GC	2.821	0.110	0.549	87.76	2.35	3.15	30.01
	FHS	2.711	0.001	0.280	91.02	2.28	3.02	27.79

		ES Properties			ES Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.968	0.614	2.444	32.40	3.08	4.62	46.04
	Normal	2.830	−0.524	0.650	32.24	2.42	3.22	23.93
	Hill	3.328	−0.026	0.603	83.50	2.53	4.10	47.07
	GC	2.222	−1.132	1.571	43.54	1.60	3.30	50.81
	FHS	3.260	−0.094	0.576	77.18	2.49	3.88	41.59
1000	HS	4.286	0.926	2.751	22.92	3.52	4.95	43.49
	Normal	2.846	−0.514	0.617	12.76	2.54	3.03	15.10
	Hill	3.380	0.020	0.446	87.04	2.75	3.84	33.29
	GC	2.135	−1.225	1.592	13.28	1.64	2.82	36.09
	FHS	3.298	−0.062	0.417	80.54	2.70	3.68	29.84

Notes to Table: We simulate T daily GARCH(1,1) losses with Student's $t(d)$ innovations (high persistence parameter configuration) and calculate VaR (top panel) and ES (bottom panel) risk measures by various methods. The two experiments correspond to estimation sample sizes equal to 500 and 1000 days. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel).

Table 4. 90% Prediction Intervals 1% VaR and ES: Low Persistence

DGP: GARCH-$t(d)$ with $\alpha = 0.10, \beta = 0.4$ and $d = 8$

		VaR Properties			VaR Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.226	0.078	0.464	82.14	2.71	3.93	38.84
	Normal	2.910	−0.238	0.321	52.02	2.63	3.19	18.08
	Hill	3.083	−0.065	0.318	84.24	2.63	3.53	28.73
	GC	3.245	0.096	0.497	84.92	2.69	3.71	32.55
	FHS	3.180	0.031	0.374	91.28	2.65	3.80	36.80
1000	HS	3.186	0.039	0.382	75.10	2.83	3.60	24.61
	Normal	2.914	−0.233	0.284	36.22	2.70	3.13	13.65
	Hill	3.091	−0.056	0.234	85.56	2.76	3.43	21.31
	GC	3.291	0.144	0.437	86.98	2.84	3.70	27.20
	FHS	3.147	0.000	0.263	90.46	2.76	3.58	25.97

		ES Properties			ES Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.867	−0.036	0.638	73.14	3.11	4.50	35.77
	Normal	3.333	−0.570	0.622	12.70	3.01	3.66	16.70
	Hill	3.857	−0.047	0.552	81.70	3.08	4.67	40.97
	GC	2.648	−1.255	1.425	40.54	2.01	3.83	46.85
	FHS	3.778	−0.125	0.533	73.52	3.03	4.41	35.33
1000	HS	3.901	−0.001	0.524	75.56	3.33	4.42	27.80
	Normal	3.337	−0.564	0.595	5.68	3.09	3.59	12.61
	Hill	3.915	0.014	0.406	86.74	3.32	4.55	31.53
	GC	2.528	−1.373	1.473	12.18	2.01	3.36	34.67
	FHS	3.820	−0.081	0.389	79.50	3.25	4.34	27.90

Notes to Table: We simulate T daily GARCH(1,1) losses with Student's $t(d)$ innovations (low persistence parameter configuration) and calculate VaR (top panel) and ES (bottom panel) risk measures by various methods. The two experiments correspond to estimation sample sizes equal to 500 and 1000 days. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel).

Table 5. 90% Prediction Intervals 1% VaR and ES: Approximately Normal Distribution

DGP: GARCH-$t\,(d)$ with $\alpha = 0.10, \beta = 0.80$ and $d = 500$

		VaR Properties			VaR Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.023	0.123	0.545	55.96	2.63	3.50	29.95
	Normal	2.889	-0.011	0.172	88.76	2.61	3.13	17.96
	Hill	2.835	-0.065	0.227	83.62	2.48	3.14	23.13
	GC	2.874	-0.026	0.209	87.06	2.54	3.16	21.68
	FHS	2.903	0.003	0.255	91.30	2.49	3.30	27.95
1000	HS	2.999	0.097	0.502	39.96	2.72	3.30	19.77
	Normal	2.895	-0.008	0.124	89.54	2.70	3.07	12.72
	Hill	2.845	-0.057	0.166	84.32	2.59	3.07	16.52
	GC	2.889	-0.014	0.150	87.92	2.65	3.11	15.65
	FHS	2.888	-0.014	0.180	90.36	2.60	3.17	19.54

		ES Properties			ES Bootstrap Intervals Properties			
T	Method	Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.447	0.129	0.647	54.48	2.92	3.83	27.26
	Normal	3.309	−0.009	0.196	88.82	2.99	3.59	17.98
	Hill	3.302	−0.016	0.318	85.56	2.79	3.74	28.82
	GC	3.416	0.098	0.489	87.92	2.79	4.31	45.91
	FHS	3.347	−0.070	0.308	80.00	2.76	3.62	26.25
1000	HS	3.480	0.160	0.608	41.90	3.09	3.80	21.25
	Normal	3.315	−0.005	0.141	89.76	3.10	3.52	12.73
	Hill	3.343	0.023	0.232	89.20	2.97	3.62	21.42
	GC	3.365	0.045	0.340	89.82	2.91	3.97	32.12
	FHS	3.274	−0.046	0.220	83.58	2.92	3.57	19.60

Notes to Table: We simulate T daily GARCH(1,1) losses with approximately normal innovations and calculate VaR (top panel) and ES (bottom panel) risk measures by various methods. The two experiments correspond to estimation sample sizes equal to 500 and 1000 days. The table reports the properties of the point estimates (left panel) of VaR and ES as well as the properties of the corresponding bootstrap intervals (right panel).

References

[1] Andreev, A., and Kanto, A. (2004). Conditional Value-at-Risk estimation using non-integer degrees of freedom of Student's t-distribution, forthcoming, *Journal of Risk*.

[2] Backus, D., Foresi, S. Li, K., and Wu, L. (1997). Accounting for Biases in Black-Scholes, manuscript, The Stern School at NYU.

[3] Baillie, R., and Bollerslev, T. (1989). The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Business and Economic Statistics* 7, 297-309.

[4] Baillie, R., and Bollerslev, T. (1992). Prediction in Dynamic Models with Time Dependent Conditional Variances, *Journal of Econometrics* 52, 91-113.

[5] Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1998). Don't Look Back, *Risk*, 11(8), 100-104.

[6] Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR without Correlations for Nonlinear Portfolios, *Journal of Futures Markets* 19, 583-602.

[7] Benson, P., and Zangari, P. (1997). A General Approach to Calculating VaR without Volatilities and Correlations, *RiskMetrics Monitor*, JP Morgan-Reuthers, Second Quarter, 19-23.

[8] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-327.

[9] Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *Review of Economics and Statistics* 69, 542-547.

[10] Christoffersen, P. (2003). *Elements of Financial Risk Management.* San Diego: Academic Press.

[11] Duan, J-C. (1994). Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance* 4, 155-167.

[12] Duffie, D., and Pan, J. (1997). An Overview of Value at Risk, *Journal of Derivatives* 4, 7-49.

[13] Engle, R. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50, 987-1007.

[14] Engle, R., and Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics* 22, 367-381.

[15] Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2000). Variance Reduction Techniques for Estimating Value-at-risk, *Management Science* 46, 1349-1364.

[16] Glasserman, P., Heidelberger, P., and Shahabuddin, P. (2002). Portfolio Value-at-Risk with Heavy-Tailed Risk Factors, *Mathematical Finance* 12, 239-270.

[17] Hill, B. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics* 3, 1163–1174.

[18] Hull, J., and White, A. (1998). Incorporating volatility updating into the historical simulation method for VAR, *Journal of Risk* 1, 5-19.

[19] Jorion, P. (1996). Risk^2: Measuring the Risk in Value-At-Risk, *Financial Analysts Journal* 52, 47-56.

[20] Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk.* Second Edition. New York: MacGraw-Hill.

[21] McNeil, A., and Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance* 7, 271-300.

[22] Pascual, L., Romo, J., and Ruiz, E. (2001). Forecasting Returns and Volatilities in GARCH Processes Using the Bootstrap, manuscript, Departamento de Estadistica y Econometria, Universidad Carlos III de Madrid.

[23] Pritsker, M. (1997). Towards Assessing the Magnitude of Value-at-Risk Errors Due to Erros in the Correlation Matrix, *Financial Engineering News*, October/November.

[24] Pritsker, M. (2001). The Hidden Dangers of Historical Simulation, Working Paper 2001-27, Federal Reserve Board, Washington, DC.

[25] Theodossiou, P. (1998). Financial Data and the Skewed Generalized t Distribution, *Management Science* 44, 1650-1661.

[26] Tsay, R. (2002). *Analysis of Financial Time Series*. New York: Wiley.

[27] Zangari, P. (1997). Streamlining the Market Risk Measurement Process, *RiskMetrics Monitor*, JP Morgan-Reuthers, First Quarter, 29-36.