

2004s-15

Estimation Risk in Financial Risk Management

Peter Christoffersen, Silvia Gonçalves

Série Scientifique
Scientific Series

Montréal
Avril 2004

© 2004 Authors. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

PARTENAIRE MAJEUR

. Ministère du développement économique et régional [MDER]

PARTENAIRES

. Alcan inc.
. Axa Canada
. Banque du Canada
. Banque Laurentienne du Canada
. Banque Nationale du Canada
. Banque Royale du Canada
. Bell Canada
. BMO Groupe Financier
. Bombardier
. Bourse de Montréal
. Caisse de dépôt et placement du Québec
. Développement des ressources humaines Canada [DRHC]
. Fédération des caisses Desjardins du Québec
. GazMétro
. Hydro-Québec
. Industrie Canada
. Ministère des Finances [MF]
. Pratt & Whitney Canada Inc.
. Raymond Chabot Grant Thornton
. Ville de Montréal

. École Polytechnique de Montréal
. HEC Montréal
. Université Concordia
. Université de Montréal
. Université du Québec à Montréal
. Université Laval
. Université McGill

ASSOCIE A :

. Institut de Finance Mathématique de Montréal (IFM²)
. Laboratoires universitaires Bell Canada
. Réseau de calcul et de modélisation mathématique [RCM²]
. Réseau de centres d'excellence MITACS (Les mathématiques des technologies de l'information et des systèmes complexes)

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

Estimation Risk in Financial Risk Management*

Peter Christoffersen[†], Silvia Gonçalves[‡]

Résumé / Abstract

La valeur-à-risque (VaR) et la mesure ES (Expected Shortfall) sont de plus en plus utilisées pour la mesure du risque d'un portefeuille, l'allocation de capital de risque et la détermination des performances. Les gestionnaires de risques financiers sont donc légitimement intéressés par la précision des techniques classiques de la valeur-à-risque et de la mesure ES. Le but de cet article est précisément d'évaluer la précision des modèles classiques et de mesurer l'importance de l'erreur d'estimation en construisant des intervalles de confiance autour des prévisions de la valeur-à-risque et de la mesure ES. Un des problèmes clés dans la construction d'intervalles de confiance appropriés provient de la dynamique de la variance conditionnelle typiquement observée pour les rendements spéculatifs. Notre article propose donc une technique de ré-échantillonnage qui tient compte de l'erreur d'estimation des paramètres des modèles dynamiques de la variance d'un portefeuille. Une analyse Monte Carlo nous montre que les méthodes généralement utilisées par les praticiens, telles que la simulation historique qui calcule le quantile empirique à l'aide d'une fenêtre mobile des rendements, génèrent des intervalles de confiance pour la valeur-à-risque à 90% qui sont trop étroits et qui contiennent seulement 20% des vraies valeurs-à-risque. D'autres méthodes qui tiennent compte correctement de la dynamique conditionnelle de la variance, telles que la simulation historique filtrée, génèrent quant à elles des intervalles de confiance de la valeur-à-risque à 90% qui contiennent près de 90% des vraies valeurs-à-risque. Les mesures ES sont généralement moins précises que les mesures de valeur-à-risque et les intervalles de confiance autour de la mesure ES sont également moins fiables.

Mots clés : gestion des risques, Bootstrap, GARCH

* We are grateful to 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 for helpful comments. FCAR, IFM2 and SSHRC provided financial support. The usual disclaimer applies.

[†] Peter 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.

[‡] Silvia 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/goncalves>.

Value-at-Risk (VaR) and Expected Shortfall (ES) are 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 and ES techniques. The purpose of this paper is exactly to assess the precision of common models and to quantify the magnitude of the estimation error by constructing confidence bands around the point VaR and ES forecasts. A key challenge in constructing proper confidence bands 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. In a Monte Carlo study we find that commonly used practitioner methods such as Historical Simulation, which calculates the empirical quantile on a moving window of returns, implies 90% VaR confidence intervals that are too narrow and that contain as few as 20% of the true VaRs. Other methods which properly account for conditional variance dynamics, such as Filtered Historical Simulation instead imply 90% VaR confidence intervals that contain close to 90% of the true VaRs. ES measures are generally less accurate than VaR measures and the confidence bands around ES are also less reliable.

Keywords: Risk management, bootstrapping, GARCH

Codes JEL : G12

1 Motivation

Value-at-Risk (VaR) and Expected Shortfall (ES) are 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 and ES 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 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 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 bands 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 (1995), Pritsker (1997), Chapell and Dowd (1999) and Dowd (2000). But a key challenge in constructing proper VaR and ES confidence bands arises from the conditional variance dynamics typically found in speculative returns. We quantify these dynamics using the celebrated GARCH models 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 bands for risk measures in this context. In a Monte Carlo study we find that commonly used practitioner VaR methods such as Historical Simulation, which calculates the empirical quantile on a moving window of returns, imply nominal 90% confidence intervals for the one-day, 1% VaR that are too narrow and that contain as few as 20% of the true VaRs. Other 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 (1999), instead imply 90% VaR confidence intervals that contain close to 90% of the true VaRs. Similarly, and importantly, we find that it is in general more difficult to construct accurate confidence bands for the ES measure. All the confidence bands from risk models we consider tend to contain the true ES less frequently than the 90% they should. The 90% confidence bands for ES are thus too narrow in general. The resampling technique we propose can be relatively easily extended to longer horizons, to multivariate risk models, and to allowing for model specification error.

In related work Figlewski (2002) computes effective coverages of VaRs with nominal coverages equal to .5, 1 and 5%. He finds that ignoring estimation error causes the effective

¹Baillie 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), Dowd (1998) Duffie and Pan (1997), and Jorion (2000) give little or no attention to the estimation error issue.

coverage to be below the nominal coverage, particularly for the smallest nominal coverage. Whereas Figlewski (2002) focuses on the average coverage of the point forecast VaR allowing for estimation risk, our focus here is to construct confidence bands around the point forecast VaR and ES.

Accurate confidence bands 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 band of 10-16% then the cautious portfolio manager should rebalance the portfolio to reduce risk. 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 bands. 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

$$L_t = \sigma_t \varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

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 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 distribution 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 exceeds VaR.

²The model can be generalized to allow for skewness following Theodossiou (1998).

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

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

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

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

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

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

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

$$ES_{T+1}^p = \sigma_{T+1} E(\varepsilon | \varepsilon > G_{1-p}^{-1}) \equiv \sigma_{T+1} c_{2,p}, \quad (5)$$

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, $ES_{T+1}^p = \sigma_{T+1} \frac{\phi(\Phi_{1-p}^{-1})}{p}$. When the innovation distribution is not the standard normal, this formula does not apply. However, we can still express the ES as σ_{T+1} times a constant $c_{2,p} \equiv E(\varepsilon | \varepsilon > G_{1-p}^{-1})$ whose value is a function of G and p . In our simulations below, where G is a standardized Student's $t(d)$ distribution, we computed $c_{2,p}$ by Monte Carlo simulation using 100,000 simulations from the appropriate distribution G .

In practice, we cannot compute the true values of VaR_{T+1}^p and ES_{T+1}^p , since they depend on the characteristics of the data generating process (i.e. they depend on G and on the conditional variance model σ_{T+1}^2). 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_{T+1}^p is given by

$$HS-VaR_{T+1}^p = Q_{1-p}(\{L_t\}),$$

where $Q_{1-p}(\{L_t\})$ denotes the $(1-p)$ -th empirical quantile of the losses data $\{L_t\}_{t=1}^T$. The HS estimate of ES_{T+1}^p is given by

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

where $\#(L_t > HS-VaR_{T+1}^p)$ denotes the number of observations of $\{L_t\}_{t=1}^T$ that are above the HS estimate of 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_{T+1}^p and ES_{T+1}^p can be obtained in three steps:

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

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

Given the QML estimates $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$, 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

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

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

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

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_{T+1}^p and ES_{T+1}^p as

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

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: fully parametric methods and nonparametric methods.

2.2 Parametric Methods

These methods postulate a distribution G and use it to compute $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$. We consider two leading cases: one based on the (erroneous) assumption of conditional normality and another based on the (true) assumption of Student's $t(d)$ distribution.

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 :

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

where

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

with Φ_{1-p}^{-1} the $(1-p)$ -th quantile of a standard normal distribution. We will call this the “Normal” method.

Student Conditional Distribution

Instead, imposing the true Student's t distribution implies the following estimates of the risk measures:

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

where

$$\begin{aligned} \hat{c}_{1,p}^t &= \sqrt{(d-2)/dt_{d,1-p}^{-1}} \\ \hat{c}_{2,p}^t &= E\left(\varepsilon | \varepsilon > \sqrt{(d-2)/dt_{d,1-p}^{-1}}\right), \end{aligned}$$

with $t_{d,1-p}^{-1}$ the $(1-p)$ -th quantile of the Student's t distribution with d degrees of freedom and $\hat{c}_{2,p}^t$ the truncated expectation of ε , a random variable following a standardized $t(d)$ distribution. We compute the truncated expectation by simulation, as explained above. This will be called the “Student” method.

Given the distributional assumption on G , no estimation is involved in computing $c_{1,p}$ and $c_{2,p}$. Thus, obtaining a prediction interval for the risk measures is equivalent to obtaining a prediction interval for $\hat{\sigma}_{T+1}$. The main disadvantage of these methods is the parametric assumption on the standardized losses. The “Normal” method imposes conditional normality, which does not hold for real data. Instead, the “Student” method imposes the (true) conditional Student's $t(d)$ distribution and is therefore unrealistic. Even if standardized losses follow a Student t distribution, in practice the number of degrees of freedom d needs to be estimated, thus introducing some estimation error which is not taken into account by this method. We consider both methods 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

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

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 $t = 2, \dots, T$). 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

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

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:

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

where

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

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 :

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

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

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

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

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

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

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

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

and

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

where $\bar{\varepsilon} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t$. 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 :

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

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 $\{L_t^* : t = 1, \dots, T\}$ 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:

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

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

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, $\{HS-VaR_{T+1}^{*p(i)} : i = 1, \dots, B\}$ 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

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

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

Following the Historical Simulation approach, this naive bootstrap method is completely nonparametric, 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 bands 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

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

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

$$\sigma_{T+1}^2 = \omega + \alpha L_T^2 + \beta \sigma_T^2. \tag{8}$$

Given (6) and (7), there are two sources⁴ of risk associated with predicting VaR_{T+1}^p and ES_{T+1}^p 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_{T+1}^2 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_T^2 as a function of past losses as follows:

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

Replacing ω, α and β with their MLE estimates yields

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

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

The presence of estimation risk in computing VaR_{T+1}^p and ES_{T+1}^p 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_{T+1}^p and ES_{T+1}^p 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 $\hat{\varepsilon}_t - \bar{\varepsilon}$, where $\hat{\varepsilon}_t = \frac{L_t}{\hat{\sigma}_t}$, $t = 1, \dots, T$. Let \hat{G}_T denote the empirical distribution function of $\hat{\varepsilon}_t$.

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

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

where $\varepsilon_t^* \sim \text{i.i.d.} \hat{G}_T$ 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

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

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

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 parametric methods (Normal and Student), we simply set

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

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 $\hat{\sigma}_1^{*2} = \hat{\omega}_1^2$. 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 $\hat{\sigma}_T^2$ is computed in (9). Notice however that $\hat{\sigma}_T^{*2}$ is conditional on the observed past observations on the losses $\{L_t : t = 1, \dots, T\}$, 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.

Depending on the estimation method, step 4 computes a bootstrap estimate of the risk measure that may or may not take into account the estimation risk in computing the constants $c_{1,p}$ and $c_{2,p}$. In particular, contrary to the nonparametric methods, the Normal and Student methods do not account for this estimation risk as they make a parametric assumption on G . 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 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 bands around the risk measures. This section provides quantitative evidence on these two issues through a Monte Carlo study.

We will consider four versions of the GARCH- $t(d)$ data generating process (DGP) below. In each version we set $\alpha = .10$ and $\omega = (20^2/252) * (1 - \alpha - \beta)$. 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 1,000 total estimation sample points respectively. In each case, we choose 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 1,000 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 3% is reasonable for VaR and 1% is reasonable for 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. In any case the results are quite similar across DGPs and sample sizes but notably different across VaR and ES models, which in itself is an important and useful finding.

Tables 1-4 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, $T = \{500, 1000\}$.

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

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.

4.1 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.

4.1.1 The Benchmark Case

The top panel of Table 1 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 quite close to the benchmark Student model, which assumes a known degree of freedom equal to 8 and is therefore not a feasible alternative. The RMSE of the Normal is also low but, as mentioned before, displays considerable bias.

Increasing the sample size to 1,000 in the second panel of Table 1 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 compared with the hypothetical benchmark Student.

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 than the hypothetical Student model. 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.

4.1.2 The High Persistence Case

The top half of Table 2 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 1 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 2 reports results for ES using the high volatility persistence DGP. We find that the results are very close to the bottom half of Table 1 for the conditional models but not for HS. This finding matches the results for VaR reported in the top halves of Tables 1 and 2 respectively. As before, the bias and RMSEs of the HS model are very large, and for the ES the GC model again performs poorly.

4.1.3 The Low Persistence Case

In the top half of Table 3 we consider the VaR case of low volatility persistence which makes the returns close to independent over time. 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.

4.1.4 The Conditional Normal Case

The top half of Table 4 contains VaR results for the conditionally normal GARCH DGP. Comparing with Table 1 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 4 considers the ES risk measure. Comparing the bottom of Table 4 with the bottom of Table 1 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 4. This is sensible from the perspective that under conditional normality the ES does not contribute information over and beyond the VaR.

4.2 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 bands and the average width of the confidence band as a percentage of the true VaR point forecast.

4.2.1 The Benchmark Case

Turning back to Table 1 and looking at the top panel, we remark that the historical simulation VaR (HS) bands (calculated from the i.i.d. bootstrap) have a terribly low effective coverage for a promised nominal coverage of 90%. Furthermore, the confidence bands are on average very wide. The HS method ignores the dynamics in the DGP which is costly both in terms of coverage and width.⁶

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 benchmark Student model suffers from slight under-coverage but is very narrow. Notice thus that the Student model is not perfect in finite samples. This is because the bootstrap is not able to perfectly capture the GARCH estimation parameter uncertainty for this sample size. Not surprisingly though the Student model with known degree of freedom performs very well overall. The Hill model has lower coverage than the Student and has wider intervals. The GC models has even lower coverage than the Hill and has wider intervals. Finally, the FHS model has slight over-coverage, which is arguably to be preferred to under-coverage, but it also has a fairly wide average coverage interval.

In the second panel of Table 1 we increase the risk manager's sample size to 1,000 past return observations in each simulation. Comparing with the top panel in Table 1 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 Student model has around the same coverage and better width compared with the 500 observation case. The estimation error in the GARCH parameters is less of an issue now, which brings the Student model closer to the truth. 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 1 reports results for the bootstrap prediction intervals from the different ES models. We notice the following: The Historical Simulation ES bands (calculated from the i.i.d. bootstrap) have a low effective coverage for a promised nominal coverage of 90%. Furthermore the confidence bands on average are quite wide. The HS results for ES are roughly comparable with those for VaR in Table 1. 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 Student model suffers from slight under-coverage but is very narrow. The Hill model has the best coverage of the feasible models but is quite wide. The GC model has very low coverage and quite wide intervals. Finally, the FHS model has considerable

⁶We 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 bootstrap is shown here because it is arguably most in line with the model-free spirit of the HS model.

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 1, we see that the infeasible Student has the best coverage followed by the Hill and then 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.

4.2.2 The High Persistence Case

The top right hand side of Table 2 reports VaR interval results from a return generating process with relatively high persistence. Comparing panel for panel with the benchmark process in Table 1, we notice that the HS model has worse coverage and worse width, whereas the Normal model has better coverage. The Student model has slightly worse coverage. The GC model has similar coverage but wider intervals. The FHS has roughly the same coverage under high persistence but the widths 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 2 reports ES results. Comparing the VaR and ES results in Table 2 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 1 reveals that the HS model has worse coverage and worse width. The Normal model has very poor coverage still. The Student model has slightly worse coverage and the width is worse. 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.

4.2.3 The Low Persistence Case

The top right hand side of Table 3 reports VaR results from returns with low variance persistence. Not surprisingly the results are reversed from Table 2, 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 Student model has near perfect coverage and 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 3 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 1.

4.2.4 The Conditional Normal Case

In Table 4 we generate returns which are close to conditionally normally distributed. Comparing the VaR panels in Table 4 with the corresponding panels in Table 1, 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 Student model has better coverage and better width, probably because the return data generating process is to some degree easier to capture. 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 4 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.

4.3 Summary of Results

Based on the results in Tables 1-4, 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 bands. 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 of the feasible models were much higher in percent of the benchmark Student model 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 bands 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 Summary and Directions for Future Work

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 bands 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 bands which have correct coverage but which are also quite wide. VaR models based on parametric assumptions 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. These findings have important implications for the choice of risk model and risk measure.

The directions for future work are several.

- So far we have studied the effects of estimation risk at the portfolio level only (See Benson and Zangari, 1997, Engle and Manganelli, 2000, 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 probably even more important in the multivariate case.
- We only consider one-day ahead VaR and ES in this paper. Extending the horizon of interest would be useful. The interesting thing about the 1-step ahead is that here ONLY estimation risk matters (given the true model).
- We want to extend the bootstrap methods described here to account for model risk.
- Risk models based on realized volatility models rather than GARCH models may allow for the calculation of VaR and ES prediction intervals from asymptotic theory (See Andersen, Bollerslev, Diebold and Labys, 2003). We plan to explore these types of models in future work.

We round off by noting 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 Truncation Points
 The Total Sample Consists of 500 Daily Loss Observations

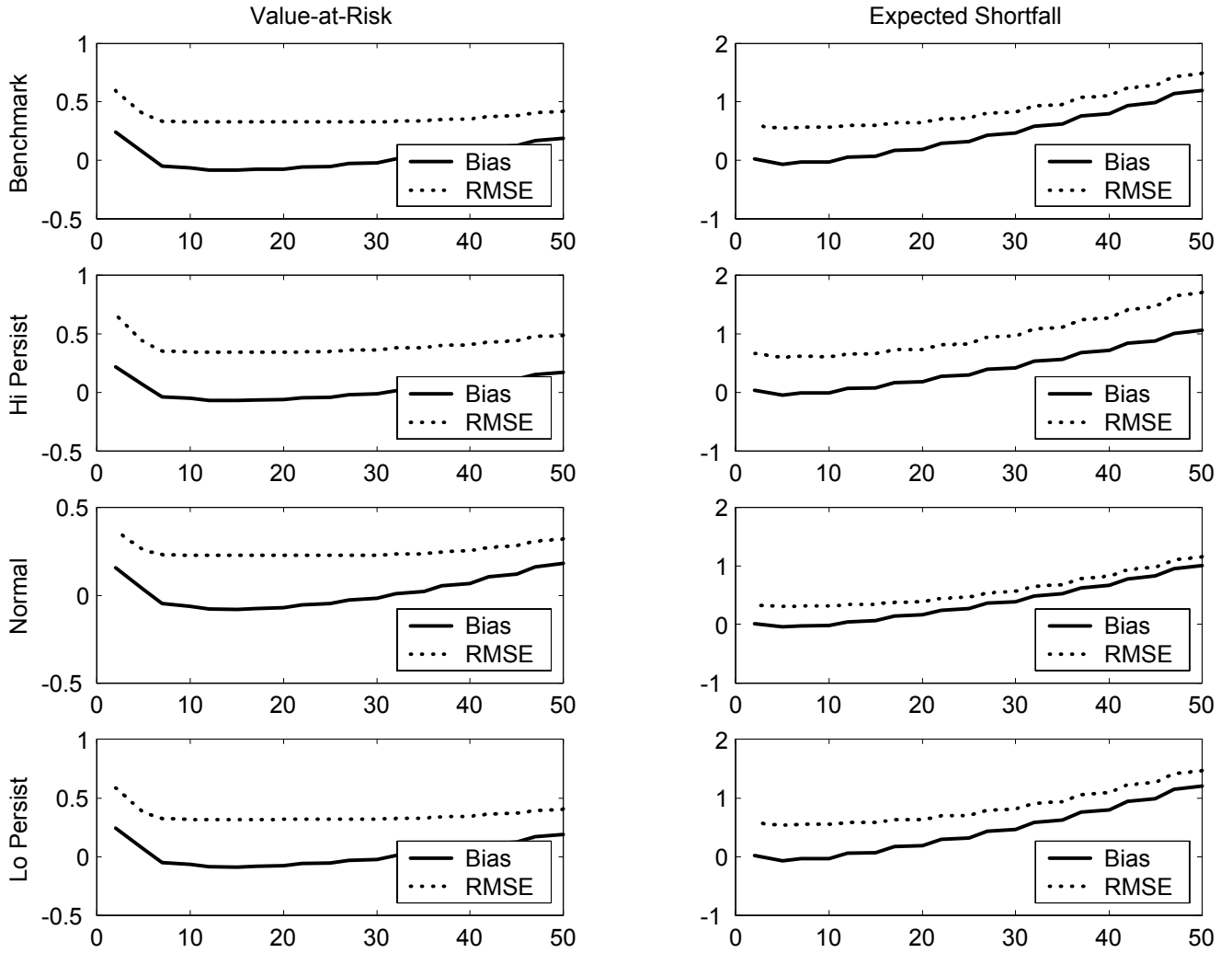


Figure 2: RMSE and Bias of Hill Estimator for Various Truncation Points
 The Total Sample Consists of 1,000 Daily Loss Observations

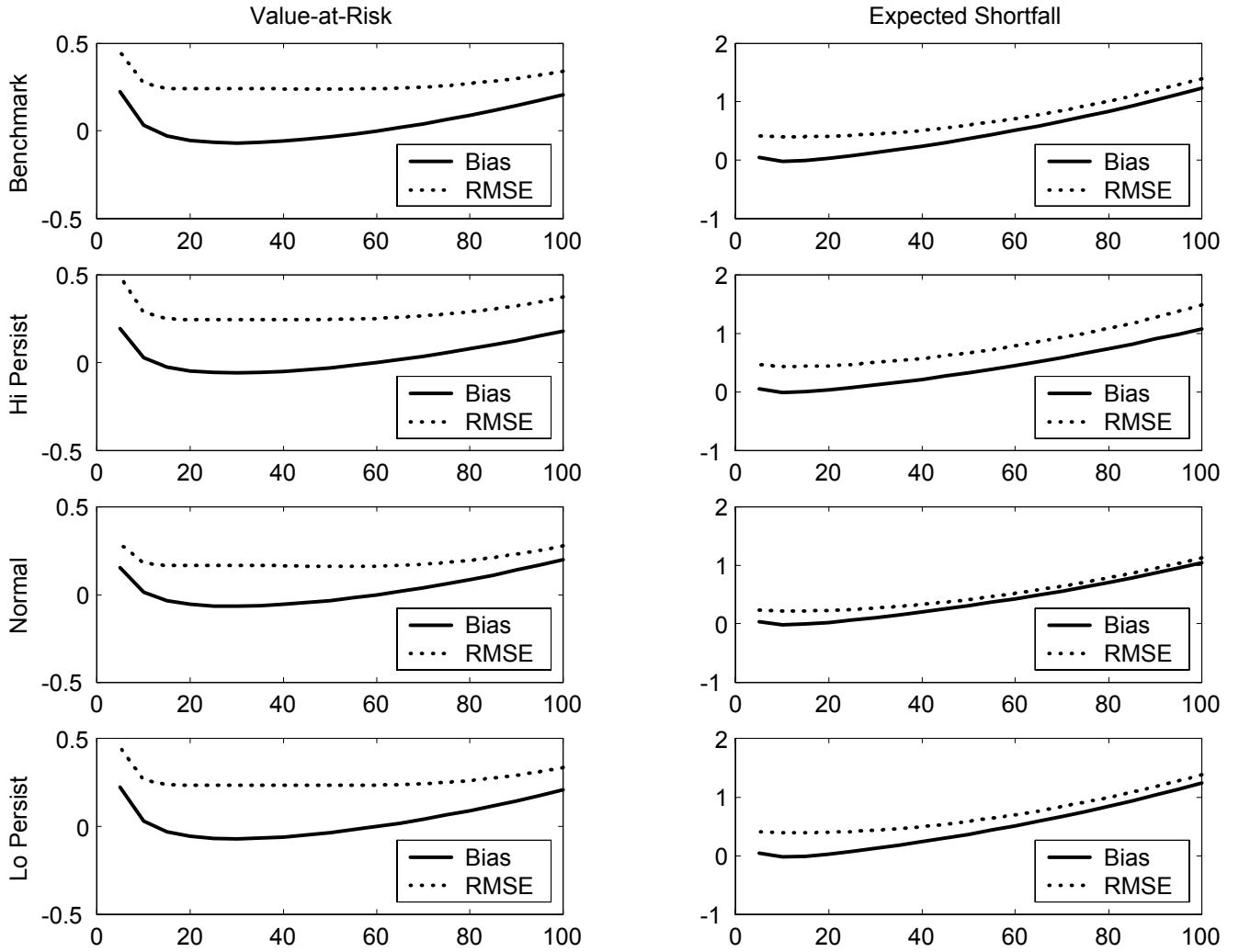


Table 1. 90% Prediction Intervals for 1% VaR and ES: Benchmark Case

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	0.176	0.747	60.44	2.73	4.03	41.80
	Normal	2.866	-0.240	0.331	59.66	2.53	3.19	21.20
	Student	3.090	-0.016	0.243	88.78	2.73	3.44	22.83
	Hill	3.053	-0.053	0.327	86.04	2.56	3.53	31.19
	GC	3.194	0.088	0.493	85.10	2.60	3.67	34.41
	FHS	3.138	0.032	0.383	90.72	2.57	3.77	38.59
1000	HS	3.239	0.133	0.671	47.52	2.85	3.69	27.11
	Normal	2.872	-0.234	0.289	41.74	2.63	3.09	14.85
	Student	3.097	-0.009	0.179	88.60	2.83	3.33	16.14
	Hill	3.071	-0.035	0.237	86.80	2.71	3.40	22.27
	GC	3.245	0.139	0.435	87.08	2.77	3.63	27.75
	FHS	3.106	0.000	0.268	90.48	2.70	3.52	26.46
T	Method	ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.966	0.131	0.978	60.54	3.15	4.61	38.06
	Normal	3.283	-0.552	0.617	20.70	2.90	3.64	19.29
	Student	3.816	-0.019	0.299	88.78	3.37	4.24	22.68
	Hill	3.765	-0.071	0.544	77.12	2.97	4.45	47.56
	GC	2.609	-1.226	1.421	42.08	1.98	3.79	47.18
	FHS	3.718	-0.107	0.535	75.42	2.95	4.36	36.75
1000	HS	4.020	0.185	0.894	52.24	3.40	4.57	30.58
	Normal	3.289	-0.546	0.586	7.32	3.01	3.54	13.85
	Student	3.824	-0.011	0.221	88.60	3.50	4.11	15.94
	Hill	3.815	-0.020	0.397	83.16	3.22	4.36	29.95
	GC	2.490	-1.345	1.469	13.52	1.98	3.32	35.02
	FHS	3.771	-0.064	0.391	80.52	3.18	4.27	28.49

Table 2. 90% Prediction Intervals 1% VaR and ES: High Persistence

DGP: GARCH- $t(d)$ with $\alpha = 0.10, \beta = 0.89$ and $d = 8$								
T	Method	VaR Properties			VaR Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.242	0.532	1.958	32.82	2.61	4.04	52.52
	Normal	2.471	-0.239	0.341	60.60	2.13	2.83	25.71
	Student	2.664	-0.046	0.239	86.68	2.30	3.05	27.54
	Hill	2.662	-0.048	0.349	85.22	2.17	3.14	35.62
	GC	2.771	0.061	0.556	85.72	2.20	3.33	41.50
	FHS	2.738	0.028	0.418	90.52	2.19	3.35	42.60
1000	HS	3.330	0.620	2.064	20.78	2.85	3.92	40.06
	Normal	2.485	-0.225	0.299	45.84	2.23	2.67	16.47
	Student	2.679	-0.031	0.178	88.44	2.41	2.88	17.60
	Hill	2.681	-0.029	0.248	87.22	2.31	2.94	23.59
	GC	2.821	0.111	0.549	87.54	2.37	3.18	30.33
	FHS	2.711	0.001	0.279	90.86	2.30	3.05	28.08
T	Method	ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.967	0.620	2.440	32.34	3.07	4.63	46.40
	Normal	2.829	-0.518	0.635	33.42	2.44	3.24	23.79
	Student	3.289	-0.058	0.296	86.68	2.84	3.77	27.66
	Hill	3.293	-0.048	0.597	78.42	2.51	3.97	43.95
	GC	2.222	-1.125	1.557	44.12	1.61	3.32	50.86
	FHS	3.259	-0.088	0.576	77.10	2.52	3.92	41.64
1000	HS	4.286	0.939	2.751	22.00	3.54	4.98	43.66
	Normal	2.846	-0.501	0.602	14.02	2.55	3.06	15.46
	Student	3.309	-0.038	0.219	88.44	2.98	3.56	17.59
	Hill	3.337	-0.010	0.435	84.46	2.75	3.80	31.72
	GC	2.135	-1.212	1.578	13.98	1.64	2.84	36.39
	FHS	3.298	-0.049	0.417	82.12	2.72	3.71	30.02

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

DGP: GARCH- $t(d)$ with $\alpha = 0.10, \beta = 0.4$ and $d = 8$								
T	Method	VaR Properties			VaR Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.226	0.079	0.464	82.36	2.72	3.94	38.79
	Normal	2.910	-0.237	0.321	52.02	2.63	3.20	18.12
	Student	3.137	-0.010	0.232	90.60	2.84	3.45	19.40
	Hill	3.093	-0.054	0.318	85.96	2.64	3.56	29.25
	GC	3.245	0.098	0.497	85.06	2.69	3.72	32.75
	FHS	3.179	0.032	0.374	91.48	2.65	3.81	36.88
1000	HS	3.186	0.039	0.382	75.10	2.83	3.60	24.47
	Normal	2.914	-0.233	0.284	36.22	2.70	3.13	13.66
	Student	3.142	-0.005	0.173	90.10	2.91	3.38	14.93
	Hill	3.112	-0.035	0.233	87.50	2.77	3.46	21.93
	GC	3.291	0.144	0.437	86.98	2.84	3.70	27.33
	FHS	3.147	0.000	0.263	90.46	2.76	3.58	26.06
T	Method	ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.867	-0.019	0.636	73.52	3.11	4.51	36.05
	Normal	3.333	-0.553	0.608	14.58	3.01	3.67	17.00
	Student	3.875	-0.011	0.286	90.60	3.50	4.26	19.57
	Hill	3.816	-0.072	0.537	76.04	3.06	4.52	37.66
	GC	2.648	-1.238	1.412	41.18	2.01	3.84	47.12
	FHS	3.778	-0.108	0.529	73.92	3.03	4.41	35.54
1000	HS	3.901	0.015	0.524	76.18	3.33	4.42	28.04
	Normal	3.337	-0.549	0.581	6.20	3.09	3.59	12.86
	Student	3.879	-0.007	0.214	90.10	3.60	4.17	14.66
	Hill	3.865	-0.021	0.392	82.88	3.29	4.44	29.55
	GC	2.528	-1.358	1.458	12.82	2.01	3.36	34.73
	FHS	3.820	-0.066	0.386	80.36	3.25	4.34	28.04

Table 4. 90% Prediction Intervals 1% VaR and ES: Normal Distribution

DGP: GARCH- $t(d)$ with $\alpha = 0.10, \beta = 0.80$ and $d = 500$								
T	Method	VaR Properties			VaR Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % VaR
500	HS	3.023	0.122	0.545	55.14	2.63	3.50	30.08
	Normal	2.889	-0.012	0.172	88.90	2.61	3.13	17.98
	Student	2.893	-0.008	0.172	88.90	2.61	3.13	17.98
	Hill	2.850	-0.051	0.227	86.06	2.48	3.17	23.86
	GC	2.874	-0.027	0.209	86.94	2.53	3.16	21.78
	FHS	2.903	0.002	0.255	90.88	2.49	3.30	28.01
1000	HS	2.999	0.098	0.502	39.96	2.72	3.30	19.98
	Normal	2.895	-0.006	0.123	89.54	2.70	3.07	12.75
	Student	2.898	-0.003	0.123	89.76	2.71	3.08	12.75
	Hill	2.869	-0.032	0.163	87.46	2.61	3.10	16.88
	GC	2.889	-0.012	0.149	87.92	2.65	3.11	15.85
	FHS	2.888	-0.013	0.179	90.36	2.60	3.17	19.64
T	Method	ES Properties			ES Bootstrap Intervals Properties			
		Average	Bias	RMSE	Coverage Rate	Lower Limit	Upper Limit	Width % ES
500	HS	3.447	0.127	0.647	54.44	2.92	3.83	27.5
	Normal	3.309	-0.011	0.196	88.86	2.99	3.58	17.83
	Student	3.309	-0.011	0.196	88.90	2.99	3.58	17.83
	Hill	3.271	-0.048	0.306	80.98	2.78	3.66	26.50
	GC	3.416	0.096	0.489	88.10	2.78	4.30	45.93
	FHS	3.347	-0.073	0.308	79.70	2.75	3.62	26.29
1000	HS	3.480	0.160	0.608	41.98	3.09	3.80	21.37
	Normal	3.315	-0.005	0.141	89.68	3.10	3.52	12.64
	Student	3.316	-0.004	0.141	89.76	3.10	3.52	12.64
	Hill	3.304	-0.017	0.219	85.54	2.95	3.61	19.82
	GC	3.365	0.045	0.339	89.90	2.91	3.97	31.91
	FHS	3.274	-0.046	0.221	83.46	2.92	3.57	19.57

References

- [1] Andersen, T., Bollerslev, T., Diebold, F.X. and P. Labys (2003), Modeling and Forecasting Realized Volatility, *Econometrica* , 71, 529-626.
- [2] Backus, D., S. Foresi, K. Li and L. Wu (1997), Accounting for Biases in Black-Scholes, manuscript, The Stern School at NYU.
- [3] Baillie, R. and T. Bollerslev (1989), The Message in Daily Exchange Rates: A Conditional Variance Tale, *Journal of Business and Economic Statistics*, 7, 297-309.
- [4] Baillie, R. and T. Bollerslev (1992), Prediction in Dynamic Models with Time Dependent Conditional Variances, *Journal of Econometrics*, 52, 91-113.
- [5] Barone-Adesi, G., K. Giannopoulos and L. Vosper (1999), VaR without Correlations for Nonlinear Portfolios, *Journal of Futures Markets*, 19, 583-602.
- [6] Benson, P. and P. Zangari (1997), A General Approach to Calculating VaR without Volatilities and Correlations, *RiskMetrics Monitor*, JP Morgan-Reuthers, Second Quarter, 19-23.
- [7] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- [8] Bollerslev, T. (1987), A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return,” *Review of Economics and Statistics*, 69, 542-547.
- [9] Chapell, D. and K. Dowd (1999), Confidence Intervals for VaR, *Financial Engineering News*, March.
- [10] Christoffersen, P. (2003), *Elements of Financial Risk Management*. San Diego: Academic Press.
- [11] Dowd. K. (1998), *Beyond Value at Risk: The New Science of Risk Management*. Chichester and New York: Wiley and Sons.
- [12] Dowd, K. (2000), Assessing VaR Accuracy, *Derivatives Quarterly*, 6, 61-63.
- [13] Duan, J-C. (1994), Maximum Likelihood Estimation Using Price Data of the Derivative Contract, *Mathematical Finance*, 4, 155-167.
- [14] Duffie, D. and J. Pan (1997), An Overview of Value at Risk, *Journal of Derivatives*, 4, 7-49.
- [15] Engle, R. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, 987-1007.
- [16] Engle, R. and S. Manganelli (2000), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, manuscript, NYU Stern and ECB.

- [17] Figlewski, S. (2002), Estimation Error in the Assessment of Financial Risk Exposure, manuscript, the Stern School of Business at New York University.
- [18] Glasserman, P., P. Heidelberger, and P. Shahabuddin (2000), Variance Reduction Techniques for Estimating Value-at-risk, *Management Science*, 46, 1349-1364.
- [19] Glasserman, P., P. Heidelberger, and P. Shahabuddin (2002), Portfolio Value-at-Risk with Heavy-Tailed Risk Factors, *Mathematical Finance*, 12, 239-270.
- [20] Hill, B. (1975), A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics*, 3, 1163–1174.
- [21] Hull, J. and A. White (1998), Incorporating volatility updating into the historical simulation method for VAR,” *Journal of Risk*, 1, Fall.
- [22] Jorion, P. (1996), Risk²: Measuring the Risk in Value-At-Risk, *Financial Analysts Journal*, 52, 47-56.
- [23] Jorion, P. (2000), *Value at Risk: The New Benchmark for Managing Financial Risk*. Second Edition. New York: MacGraw-Hill.
- [24] McNeil, A. and R. Frey (2000), Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance*, 7, 271-300.
- [25] Pascual, L., J. Romo, and E. Ruiz (2001), Forecasting Returns and Volatilities in GARCH Processes Using the Bootstrap, manuscript, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.
- [26] Pritsker, M. (1997), Towards Assessing the Magnitude of Value-at-Risk Errors Due to Errors in the Correlation Matrix, *Financial Engineering News*, October/November.
- [27] Pritsker, M. (2001), The Hidden Dangers of Historical Simulation, Working Paper 2001-27, Federal Reserve Board, Washington, DC.
- [28] Theodossiou, P. (1998), Financial Data and the Skewed Generalized t Distribution, *Management Science*, 44, 1650-1661.
- [29] Zangari, P. (1997), Streamlining the Market Risk Measurement Process, *RiskMetrics Monitor*, JP Morgan-Reuthers, First Quarter, 29-36.