Uncertainty of Multiple Period Risk Measures*

Carl Lönnbark
Department of Economics, Umeå University
SE-901 87 Umeå, Sweden
E-mail: carl.lonnbark@econ.umu.se

Abstract

In general, the properties of the conditional distribution of multiple period returns do not follow easily from the one-period data generating process. This renders computation of Value-at-Risk and Expected Shortfall for multiple period returns a non-trivial task. In this paper we consider some approximation approaches to computing these measures. Based on the results of a simulation experiment we conclude that among the studied analytical approaches the one based on approximating the distribution of the multiple period shocks by a skew-t was the best. It was almost as good as the simulation based alternative. We also found that the uncertainty due to the estimation risk can be quite accurately estimated employing the delta method. In an empirical illustration we computed five day VaR's for the S&P 500 index. The approaches performed about equally well.

Key Words: Asymmetry, Estimation Error, Finance, GJR-GARCH, Prediction, Risk Management.
JEL Classification: C16, C46, C52, C53, C63, G10.

Umeå Economic Studies 768, 2009

*The financial support from the Wallander-Hedelius Foundation is gratefully acknowledged. Kurt Brännäs and Johan Lyhagen are thanked for many valuable comments and suggestions.

1 Introduction

The focus of this paper is on predicting the risk for multiple period asset returns. An important example when this is of interest is for the market-risk charge of the Basel Committee on Banking Supervision (Basel), that is based on an horizon of 10 trading days. The market risk is defined as the risk of adverse movements in the prices of the assets in the portfolio and the measure underlying the market risk charge is the Value at Risk (VaR) (defined below). Basel allows financial institutions to compute the 10 day VaR by multiplying the one day VaR by the square root of 10. However, it is well known (e.g., Diebold, Hickman, Inoue, and Schuermann, 1997) that this approach (Root-k) may give very erroneous VaR's and alternative approaches are thus called for.

When it comes to predicting more than one period ahead there are two approaches: The direct approach specifies a model for the relevant horizon, e.g., 10 days, directly, whereas the iterating approach iterates on a model specified for a shorter horizon, e.g., one day, to obtain the multiple period predictions. The first approach may be more robust to misspecification, while the latter may produce more efficient parameter estimates (e.g., Marcellino, Stock, and Watson, 2006; Pesaran, Pick, and Timmermann, 2009). The recommendation put forth by Diebold et al. (1997) is to use the direct approach for risk predictions. Taylor (1999, 2000) propose a regression quantile approach that may be viewed as a combination of the two. In practise, the computed risk measures are subject to estimation error. Assume for example that we wish to predict the risk of an asset for a 10 day horizon and that we have two years of daily return data. For the iterating approach we would typically specify a model for the daily returns and base the prediction on the full sample of approximately 500 observations. For the direct approach on the other hand we would have only 50 observations, which may not be enough for producing a reliable prediction. We view this as a valid concern and focus here on the iterating approach. Of course, an important underlying question that we neglect here is that of whether the properties of the return distribution can be considered predictable for a particular horizon (see Christoffersen and Diebold, 2000, for a discussion on volatility predictability).

As measures of (market) risk we consider VaR and the Expected Shortfall (ES). The VaR has become the standard measure of market risk and it is commonly employed by financial institutions and their regulators. The VaR has already received much attention in the literature (see Jorion, 2007, for a survey) and it is defined as the maximum potential loss over a given horizon that will not be exceeded with a given probability, or

$$\Pr\left\{\text{portfolio loss} \geq VaR^{1-\alpha}\right\} = \alpha.$$

The probability $1-\alpha$ is commonly referred to as the confidence level of the VaR. The attractive feature of the VaR is that it summarizes the properties of the return distribution into an easily interpreted number. However, it does not tell the risk manager anything about the size of the loss when disaster strikes. A measure that does exactly that is the ES. It is defined as

$$ES^{1-\alpha} = E\left(\text{portfolio loss} \mid \text{portfolio loss} \geq VaR^{1-\alpha}\right).$$

Suppose now that the risk manager wants to assess the k-period risk of the portfolio and decides to employ the iterating approach within the popular GARCH framework of Engle (1982) and Bollerslev (1986). A problem that arises is then that the properties of the multiple period return distribution may not follow easily from the one-period model. For example, even though the multiple period conditional variance implied from a one-period GARCH model with normal innovations is tractable, less so is the distribution of the corresponding innovation (Boudoukh, Richardson, and Whitelaw, 1997). Brummelhuis and Guégan (2005) provide a theoretical discussion on the matter. In particular, they show that the Root-k rule may fail severely for small values on α (see also Brummelhuis and Kaufmann, 2007).

Two alternative approaches are to compute the measures either by simulation (cf. McNeil and Frey, 2000) or to consider some analytic approximation. The former computes the measures as empirical counterparts for multiple period returns simulated from the one-period model. Assuming that the true parameters of the one-period model are known, the simulation approach can give measures arbitrarily close to the true ones. We will discuss two analytical approximations. The first one uses a Gram-Charlier expansion of the conditional density of the multiple period returns. The second one was proposed by Wong and So (2003, 2007) in related studies. It consists of specifying a conditional distribution for the multiple period returns and of obtaining the parameters of that distribution by matching its moments to the theoretical ones implied by the one-period model. The obvious benefit of using analytic approximations is that they require less computer time. In Cotter (2007) an approach based on extreme value theory is proposed. It performed poorly in simulations, though, and we do not consider it here.

As noted above, an additional source of uncertainty of the risk predictors arises from the fact that the parameters of the underlying model are unknown, which gives rise to estimation error. We also pay attention to this source of error, which is not done in Wong and So (2003). Note that this uncertainty comes in in two places for the simulation based predictor. Not only in estimating the parameters of the underlying model, but also in the second step when the measures are obtained from the simulated returns.

The uncertainty in risk prediction should be of concern to risk managers. Surprisingly little work has been done on it though and the predictions are often reported as if they were true constants. For example Lan, Hu, and Johnson (2007) report that the research on the uncertainty of VaR only amounts to about 2.5 percent of the VaR literature. One study that recognizes that VaR and ES predictors are subject to uncertainty is Christoffersen and Gonçalves (2005), who use resampling techniques to study the uncertainty of VaR and ES predictors in a GARCH framework. The obvious disadvantage of their method is that it is time consuming since it amounts to repeated estimation of a possibly complicated model. Analytical expressions (when sufficiently accurate) to quantify the uncertainty are obviously preferred. For this purpose Chan, Deng, Peng, and Xia (2007) and others consider the conventional delta method, which is done here as well.

The paper is organized as follows. In Section 2 the approaches to computing the multiple period VaR and ES are introduced. In Section 3 we discuss how to quantify the uncertainty due to the estimation error. An example is given in Section 4, where analytical results are given for the asymmetric GARCH (GJR-GARCH) model of Glosten, Jagannathan, and Runkle (1993). Section 5 contains a simulation study of the predictors obtained from the GJR-GARCH. In Section 6 an empirical illustration for the S&P 500 index is included. The final section concludes.

2 Multiple period VaR and ES

Denote by $\mathbf{w} = (w_1, ..., w_M)'$ the time invariant vector of portfolio weights between T and $T+k$. The log-return (return) between T and $T+k$ for the portfolio is approximately $\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'(\mathbf{y}_{T+1} + ... + \mathbf{y}_{T+k})$, where $\mathbf{y}_{T+l} = (y_{1,T+l}, ..., y_{M,T+l})'$, $l = 1, ..., k$, is a M-dimensional vector of one-period returns. Denote by Ψ_T the information set at time T and let the vector $\boldsymbol{\theta}$ contain the parameters governing the data generating process with $\boldsymbol{\theta}_0$ denoting true values. In practise, the information available to the risk manager is some realization of the partition, $\mathcal{F}_{t_0,T} = (\mathbf{x}_{t_0}, ..., \mathbf{x}_T)$, of Ψ_T and where $\mathbf{x}_t$, $t = t_0, ..., T$, typically contains past asset returns. A realization of the random partition, $\mathcal{F}_{t_0,T}$, is denoted by $\mathcal{F}_{t_0,T}$. Denote by $f_{T,k}(\cdot)$ and $F_{T,k}(\cdot)$ the density function (*pdf*) and distribution function (*cdf*) of $\mathbf{w}'\mathbf{Y}_{T,k}$ conditional on Ψ_T. Also, let $\boldsymbol{\mu}_{T,k}$ be the vector valued conditional mean function and $\mathbf{H}_{T,k}$ the matrix valued conditional variance-covariance function of $\mathbf{Y}_{T,k}$. We will assume that it is possible to obtain the exact forms of these conditional moments for all k.

Now, assume that the vector process, $\mathbf{y}_t$, of the asset returns started in the infinite past and that it is generated in discrete time up through, at least, $T + k$ by

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{H}_t^* \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{\varepsilon}_t$ has mean $\mathbf{0}$ and the identity matrix, $\mathbf{I}$, as its variance-covariance matrix conditional on the information available at $t-1$. Then, $\boldsymbol{\mu}_t$ is the conditional mean of $\mathbf{y}_t$, whereas $\mathbf{H}_t = \mathbf{H}_t^* \mathbf{H}_t^{*\prime}$ is the conditional variance-covariance matrix.

The conditional VaR for the period T to $T+k$ portfolio return satisfies

$$P\left(\mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha} | \Psi_T\right) = \int_{-\infty}^{-VaR_{T,k}^{1-\alpha}} f_{T,k}(y)\, dy = \alpha. \tag{2}$$

The associated conditional ES is defined as

$$\begin{aligned} ES_{T,k}^{1-\alpha} &= -E_T\left(\mathbf{w}'\mathbf{Y}_{T,k} \mid \mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha}\right) \\ &= -\frac{1}{\alpha}\int_{-\infty}^{-VaR_{T,k}^{1-\alpha}} y f_{T,k}(y)\, dy, \end{aligned} \tag{3}$$

where $E_T(\cdot)$ is shorthand for expectation conditional on Ψ_T. The minus signs in (2) and (3) stem from the convention of reporting VaR and ES as positive numbers.

For $k = 1$, $VaR_{T,1}^{1-\alpha}$ and $ES_{T,1}^{1-\alpha}$ can (in principle) be obtained directly from (1) along with a distributional assumption on $\boldsymbol{\varepsilon}_{T+1}$. Although complications may arise in this case as well we choose here to focus on the case when $k > 1$. The further issue is then one of temporal aggregation and our point of departure is that it is not possible to obtain $VaR_{T,k}^{1-\alpha}$ and $ES_{T,k}^{1-\alpha}$ analytically and that we have to resort to some approximations $\widetilde{VaR}_{T,k}^{1-\alpha}$ and $\widetilde{ES}_{T,k}^{1-\alpha}$. We consider three such approaches. One is simulation based and targets the measures directly, whereas the other two are analytical approximations and start from an approximation to a zero mean and unit variance random variable, $\varepsilon_{T,k}$.

Denote by $VaR_{T,k}^{S,1-\alpha}$ and $ES_{T,k}^{S,1-\alpha}$ the values of $\widetilde{VaR}_{T,k}^{1-\alpha}$ and $\widetilde{ES}_{T,k}^{1-\alpha}$ computed by the simulation approach. To explain the approach, we first assume that observations are available up through T. We then simulate returns $\mathbf{y}_{T+1}^r$, $\mathbf{y}_{T+2}^r, ..., \mathbf{y}_{T+k}^r$, $r = 1, ..., R$, from the model (1) and compute the k-period portfolio returns $\mathbf{w}'\mathbf{Y}_{T,k}^r$, $r = 1, ..., R$. The VaR is obtained as the αth empirical quantile of the simulated portfolio returns, or

$$VaR_{T,k}^{S,1-\alpha} = -(\mathbf{w}'\mathbf{Y}_{T,k})_{(\alpha R+1)} \tag{4}$$

where $(\mathbf{w}'\mathbf{Y}_{T,k})_{(r)}$ is the rth order statistic of simulated returns. The corresponding ES is given by

$$ES_{T,k}^{S,1-\alpha} = -\frac{\sum_{r=1}^{\alpha R+1}(\mathbf{w}'\mathbf{Y}_{T,k})_{(r)}}{\alpha R+1}. \tag{5}$$

Note that $\mathbf{w}'\mathbf{Y}_{T,k}^r$ is iid and it is well-known that the resulting estimators are consistent. Given R though, one may of course argue that more efficient related estimators based on kernel functions exist. Chen and Tang (2005) and Chen (2008) found that, for the kernel estimator proposed by Scaillet (2004), this is the case for VaR but not necessarily for ES. Note however that R is at our discretion and extra precision comes at a small marginal cost for models within a reasonable degree of complexity.

For the analytical approaches we first assume that the k-period portfolio return, $\mathbf{w}'\mathbf{Y}_{T,k}$, admits the scale-location representation

$$\mathbf{w}'\mathbf{Y}_{T,k} = \mathbf{w}'\boldsymbol{\mu}_{T,k} + \varepsilon_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{6}$$

where $\varepsilon_{T,k}$ has zero mean, unit variance, conditional third moment, $s_{T,k}$, conditional fourth moment, $k_{T,k}$, and conditional density function

$$g_{T,k}(\varepsilon) = \sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}f_{T,k}(\boldsymbol{\mu}_{T,k} + \varepsilon\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}).$$

From (6) we then have that

$$\begin{aligned} &P\left(\mathbf{w}'\mathbf{Y}_{T,k} \leq -VaR_{T,k}^{1-\alpha}|\Psi_T\right) = P\Big(\mathbf{w}'(\mathbf{Y}_{T,k} - \boldsymbol{\mu}_{T,k})/\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}} \\ &\leq (-VaR_{T,k}^{1-\alpha} - \mathbf{w}'\boldsymbol{\mu}_{T,k})/\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}|\Psi_T\Big) = P\left(\varepsilon_{T,k} \leq q_{T,k}^{\alpha}|\Psi_T\right), \end{aligned}$$

where $q_{T,k}^{\alpha}$ solves $\alpha = \int_{-\infty}^{q_{T,k}^{\alpha}} g_{T,k}(\varepsilon)\, d\varepsilon$. The conditional portfolio VaR is then given by

$$VaR_{T,k}^{1-\alpha} = -\mathbf{w}'\boldsymbol{\mu}_{T,k} - q_{T,k}^{\alpha}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}. \tag{7}$$

The conditional ES of the portfolio is

$$ES_{T,k}^{1-\alpha} = -\mathbf{w}_T'\boldsymbol{\mu}_{T,k} - e_{T,k}^{\alpha}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, \tag{8}$$

where $e^{\alpha}_{T,k} = E_T\left(\varepsilon_{T,k} \mid \varepsilon_{T,k} \leq q^{\alpha}_{T,k}\right)$. We previously assumed that it was possible to obtain the exact analytical forms of $\boldsymbol{\mu}_{T,k}$ and $\mathbf{H}_{T,k}$. The problem is then one of approximating the density $g_{T,k}(\cdot)$. Denote this approximation by $\tilde{g}_{T,k}(\cdot)$ and the associated VaR and ES are then

$$\begin{aligned} \widetilde{VaR}^{1-\alpha}_{T,k} &= -\mathbf{w}'\boldsymbol{\mu}_{T,k} - \tilde{q}^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, &(9)\\ \widetilde{ES}^{1-\alpha}_{T,k} &= -\mathbf{w}'\boldsymbol{\mu}_{T,k} - \tilde{e}^{\alpha}_{T,k}\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}, &(10)\end{aligned}$$

where $\tilde{q}^{\alpha}_{T,k}$ satisfies $\alpha = \int_{-\infty}^{\tilde{q}^{\alpha}_{T,k}} \tilde{g}_{T,k}(\varepsilon)\, d\varepsilon$ and $\tilde{e}^{\alpha}_{T,k} = \widetilde{E}_T(\varepsilon_{T,k} \mid \varepsilon_{T,k} \leq \tilde{q}^{\alpha}_{T,k})$. Note that $\widetilde{E}_T$ is the expectation operator with respect to $\tilde{g}_{T,k}(\cdot)$.

Our first analytical approximation employs an expansion of $g_{T,k}(\cdot)$ allowing for skewness and excess kurtosis. We assume that $g_{T,k}(\cdot)$ admits the Gram-Charlier Type A expansion

$$g_{T,k}(\varepsilon) = \sum_{i=0}^{\infty} c_i H_i(\varepsilon)\,\phi(\varepsilon), \tag{11}$$

where the constants, c_i, are functions of the conditional moments of $\varepsilon_{T,k}$, $H_i(\cdot)$ are the Hermite polynomials and $\phi(\cdot)$ is the standard normal *pdf*. The sum in (11) is usually truncated at a small value of i. Jondeau and Rockinger (2001) identify the versions typically adopted in the literature to be the Edgeworth expansion and the Gram-Charlier expansion. The latter is given by

$$\tilde{g}_{T,k}(\varepsilon) = [1 + \frac{s_{T,k}}{6} H_3(\varepsilon) + \frac{k_{T,k} - 3}{24} H_4(\varepsilon)]\phi(\varepsilon). \tag{12}$$

The Edgeworth expansion adds the term $s^2_{T,k} H_6(\varepsilon)/72$ to the expression inside the brackets in (12). Barton and Dennis (1952) show that the region of $(s_{T,k}, k_{T,k})$-pairs guaranteeing positive values is larger for the Gram-Charlier expansion, and for that reason, Jondeau and Rockinger (2001) focus on the latter and so do we.

The αth quantile implied by the Gram-Charlier density in is given by the Cornish-Fisher expansion (Cornish and Fisher, 1938; see also Baillie and Bollerslev, 1992, for a related use)

$$\tilde{q}^{\alpha}_{T,k} = \Phi^{-1}_{\alpha} + \frac{s_{T,k}}{6}[(\Phi^{-1}_{\alpha})^2 - 1] + \frac{k_{T,k} - 3}{24}[(\Phi^{-1}_{\alpha})^3 - 3\Phi^{-1}_{\alpha}] - \frac{s^2_{T,k}}{36}[2(\Phi^{-1}_{\alpha})^3 - 5\Phi^{-1}_{\alpha}].$$

The third, $s_{T,k}$, and the fourth, $k_{T,k}$, conditional moments of $\varepsilon_{T,k}$ are derived from the one-period model. Christoffersen and Gonçalves (2005) propose a corresponding $\tilde{e}^{\alpha}_{T,k}$. Giamouridis (2006) correctly argues that their expression is incorrect and propose

$$\tilde{e}^{\alpha}_{T,k} = -\frac{\phi_{\tilde{q}^{\alpha}_{T,k}}}{\alpha}\left\{1 + \frac{s_{T,k}}{6}\left(\tilde{q}^{\alpha}_{T,k}\right)^3 + \frac{k_{T,k} - 3}{24}\left[\left(\tilde{q}^{\alpha}_{T,k}\right)^4 - 2\left(\tilde{q}^{\alpha}_{T,k}\right)^2 - 1\right]\right\}.$$

The VaR and the ES are obtained by plugging the expressions above into (9) and (10), respectively. We denote the resulting approximations by $VaR^{GC,1-\alpha}_{T,k}$ and $ES^{GC,1-\alpha}_{T,k}$.

Alternatively, Wong and So (2003) assume a distribution for $\varepsilon_{T,k}$ and obtaining the parameters of that distribution involves matching the third and the fourth moments to the corresponding ones implied by the one-period model. We denote the resulting approximations by $VaR_{T,k}^{WS,1-\alpha}$ and $ES_{T,k}^{WS,1-\alpha}$.

For comparison we also include the Root-k approach. The k-period VaR and the ES are then simply approximated by

$$\begin{aligned} VaR_{T,k}^{Rk,1-\alpha} &= \sqrt{k}VaR_{T,1}^{1-\alpha}, \\ ES_{T,k}^{Rk,1-\alpha} &= \sqrt{k}ES_{T,1}^{1-\alpha}. \end{aligned}$$

3 Estimation error

The traditional estimator of the parameter vector, $\boldsymbol{\theta}$, in model (1) has over the years been (conditional) maximum likelihood with a normality assumption on $\boldsymbol{\varepsilon}_t$, i.e.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ L_T(\boldsymbol{\theta}) \propto -\frac{1}{2} \sum_{t=t_0+s}^{T} [\ln|\mathbf{H}_t| + (\mathbf{y}_t - \boldsymbol{\mu}_t)'\mathbf{H}_t^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t)] \right\}. \tag{13}$$

Here, s is determined by the number of lags in $\boldsymbol{\mu}_t$ and $\mathbf{H}_t$. Given some regularity conditions the estimator, $\hat{\boldsymbol{\theta}}$, is asymptotically, normally distributed with the true parameter vector, $\boldsymbol{\theta}_0$, as its mean and with variance-covariance matrix $\boldsymbol{\Sigma}$, which may be consistently estimated by $T^{-2}[\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}']$ or $\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$. As shown by Bollerslev and Wooldridge (1992) and others, the estimator (13) remains consistent and asymptotically normal even if the distribution of $\boldsymbol{\varepsilon}_t$ is non-normal. The estimator is then known as the Quasi-Maximum Likelihood (QML) estimator and we would use the robust sandwich form as the estimator of $\boldsymbol{\Sigma}$, i.e. $(\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}(\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}')(\partial^2 L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')^{-1}$.

For all four approaches the approximated risk measures are functions of the parameters $\boldsymbol{\theta}$. Therefore, the measures are not only subject to an approximation error, but also to the estimation error in $\hat{\boldsymbol{\theta}}$. In the first approach this shows up in the simulations as they are made from the model (1) under $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The other three predictors are obtained by plugging the estimator, $\hat{\boldsymbol{\theta}}$, into (9) and (10) to obtain

$$\begin{aligned} \widehat{VaR}_{T,k}^{1-\alpha} &= -\mathbf{w}'\hat{\boldsymbol{\mu}}_{T,k} - \hat{q}_{T,k}^{\alpha}\sqrt{\mathbf{w}'\hat{\mathbf{H}}_{T,k}\mathbf{w}} \\ \widehat{ES}_{T,k}^{1-\alpha} &= -\mathbf{w}'\hat{\boldsymbol{\mu}}_{T,k} - \hat{e}_{T,k}^{\alpha}\sqrt{\mathbf{w}'\hat{\mathbf{H}}_{T,k}\mathbf{w}}. \end{aligned}$$

Early attempts (see Schmidt, 1974) to quantify the effect on prediction of errors in parameters relied on the asymptotic distribution of the parameter estimator assumed to be independent of the conditioning information. In the notation set out in the beginning of Section 2 the predictors are functions of $\Psi_{t_0,T}$ both directly and indirectly through $\hat{\boldsymbol{\theta}}$. Denote this (continuous) function by $u\left(\Psi_{t_0,T}, \hat{\boldsymbol{\theta}}(\Psi_{t_0,T})\right)$. The approach then amounts to conditioning the first argument of $u(\cdot)$ on a realization $\bar{\Psi}_{t_0,T}$ and viewing randomness to arise through the random $\Psi_{t_0,T}$ in the second argument. This approach now appears to be the convention (see Kaibila and He, 2004, for a recent discussion) and is chosen in this paper as well. In a related study Hansen (2006) takes this route and shows that, for $k = 1$, $\sqrt{T^*}\left(\widehat{VaR}^{1-\alpha}_{T,1} - VaR^{1-\alpha}_{T,1}\right) \xrightarrow{d} N\left(0, T^*\sigma^2_{VaR,T,1}\right)$, where $T^* = T - (t_0 + s)$, $\sigma^2_{VaR,T,1} = \partial VaR^{1-\alpha}_{T,1}/\partial\boldsymbol{\theta}'\ \boldsymbol{\Sigma}\ \partial VaR^{1-\alpha}_{T,1}/\partial\boldsymbol{\theta}$ and where the limit is for $t_0 \to -\infty$. This approach is directly applicable to the analytical approximations approach. They are all functions of the estimator, $\hat{\boldsymbol{\theta}}$, and the information set, $\Psi_{t_0,T}$. By the same logic as above we have that

$$\sqrt{T^*}\left[u\left(\bar{\Psi}_{t_0,T}, \hat{\boldsymbol{\theta}}\right) - u\left(\bar{\Psi}_{t_0,T}, \boldsymbol{\theta}_0\right)\right] \xrightarrow{d} N\left(0, T^*\sigma^2_u\right), \tag{14}$$

where $\sigma^2_u = \partial u/\partial\boldsymbol{\theta}'\boldsymbol{\Sigma}\partial u/\partial\boldsymbol{\theta}$ and note that it is a function of $\bar{\Psi}_{t_0,T}$. Explicitely, the variance expressions for the VaR and ES approximations are, respectively, given by

$$\begin{aligned} \sigma^2_{VaR,T,k} &= \frac{\partial\widetilde{VaR}^{1-\alpha}_{T,k}}{\partial\boldsymbol{\theta}'}\ \boldsymbol{\Sigma}\ \frac{\partial\widetilde{VaR}^{1-\alpha}_{T,k}}{\partial\boldsymbol{\theta}} \\ \sigma^2_{ES,T,k} &= \frac{\partial\widetilde{ES}^{1-\alpha}_{T,k}}{\partial\boldsymbol{\theta}'}\ \boldsymbol{\Sigma}\ \frac{\partial\widetilde{ES}^{1-\alpha}_{T,k}}{\partial\boldsymbol{\theta}}, \end{aligned}$$

where $\partial\widetilde{VaR}^{1-\alpha}_{T,k}/\partial\boldsymbol{\theta} = -\ \mathbf{w}'\partial\boldsymbol{\mu}_{T,k}/\partial\boldsymbol{\theta}\ - \sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}\partial\tilde{q}^{\alpha}_{T,k}/\partial\boldsymbol{\theta}\ -\ \tilde{q}^{\alpha}_{T,k}\mathbf{w}'\partial\mathbf{H}_{T,k}/\partial\boldsymbol{\theta}\ \mathbf{w}/(2\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}})$ and $\partial\widetilde{ES}^{1-\alpha}_{T,k}/\partial\boldsymbol{\theta} = -\ \mathbf{w}'\partial\boldsymbol{\mu}_{T,k}/\partial\boldsymbol{\theta} - \sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}}\partial\tilde{e}^{\alpha}_{T,k}/\partial\boldsymbol{\theta} - \tilde{e}^{\alpha}_{T,k}\mathbf{w}'\ \partial\mathbf{H}_{T,k}/\partial\boldsymbol{\theta}\ \mathbf{w}/(2\sqrt{\mathbf{w}'\mathbf{H}_{T,k}\mathbf{w}})$. In practise, estimators of the derivatives are obtained by plugging in $\hat{\boldsymbol{\theta}}$.

Regarding the uncertainty of the simulation based predictor we first recognize that it is a two-step procedure. The first step consists of estimating the model based on the available observations, whereas the predictors in the second step are obtained based on simulated returns from the estimated model. Hence, the estimation uncertainty comes from two sources. Now, for notational convenience drop the time indices on the *pdf* and the *cdf* of the k-period portfolio return and extend the functions to $f(\cdot;\boldsymbol{\theta})$ and $F(\cdot;\boldsymbol{\theta})$ to indicate the value of the parameter. Also, let $v_{\hat{\boldsymbol{\theta}}}$ and $e_{\hat{\boldsymbol{\theta}}}$ (not to be confused with $e^{\alpha}_{T,k}$ above) denote the true $VaR^{1-\alpha}_{T,k}$ and $ES^{1-\alpha}_{T,k}$ under the parameterization $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Now, it is possible to show (see Manistre and Hancock, 2005, and references therein) that conditional on $\hat{\boldsymbol{\theta}}$

$$\hat{v}_{\hat{\boldsymbol{\theta}}} \overset{asy}{\sim} N\left(v_{\hat{\boldsymbol{\theta}}}, V^v_{\hat{\boldsymbol{\theta}}}\right) \tag{15}$$

$$\hat{e}_{\hat{\boldsymbol{\theta}}} \overset{asy}{\sim} N\left(e_{\hat{\boldsymbol{\theta}}}, V^e_{\hat{\boldsymbol{\theta}}}\right), \tag{16}$$

where $V^v_{\hat{\boldsymbol{\theta}}} = \alpha(1-\alpha)/(f(v_{\hat{\boldsymbol{\theta}}};\hat{\theta})^2 R)$ and $V^e_{\hat{\boldsymbol{\theta}}} = [V(Y_{T,k} \mid Y_{T,k} < v_{\hat{\boldsymbol{\theta}}}) + (1-\alpha)(e_{\hat{\boldsymbol{\theta}}} - v_{\hat{\boldsymbol{\theta}}})^2]/(R\alpha)$. Of course, these variances do not recognize that $\hat{\boldsymbol{\theta}}$ is random. To derive such expressions we use the variance decomposition formula and take a first order expansion around $\boldsymbol{\theta}_0$. Ignoring higher order terms we have for $\hat{v}_{\boldsymbol{\theta}}$ that

$$\begin{aligned} V(\hat{v}_{\hat{\boldsymbol{\theta}}}) &= E[V^v_{\hat{\boldsymbol{\theta}}}] + V[v_{\hat{\boldsymbol{\theta}}}] \\ &= E[V^v_{\boldsymbol{\theta}_0} + \frac{\partial V^v_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] + V[v_{\boldsymbol{\theta}_0} + \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] \\ &= V^v_{\boldsymbol{\theta}_0} + \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}'}\ \boldsymbol{\Sigma}\ \frac{\partial v_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}}, \end{aligned}$$

where the expectation and the variance are taken over $\hat{\theta}$, and where the first approximation is motivated by (15) and the second and third ones by the asymptotic properties of $\hat{\boldsymbol{\theta}}$. The corresponding expression for $\hat{e}_{\hat{\boldsymbol{\theta}}}$ is

$$V(\hat{e}_{\hat{\boldsymbol{\theta}}}) = V^e_{\boldsymbol{\theta}_0} + \frac{\partial e_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}'} \boldsymbol{\Sigma} \frac{\partial e_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}.$$

4 Approximations: An example

The discussion so far has been in a multivariate context, i.e. the conditional mean and the conditional covariance function appeared explicitly in the expression for the portfolio returns. We drop that explicitness here and assume that one-period returns are generated by

$$\begin{aligned} y_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1} + \gamma \mathbf{1}\left(y_{t-1} < 0\right) y_{t-1}^2, \end{aligned} \tag{17}$$

where ε_t is standard normally distributed and $\mathbf{1}(\cdot)$ is the indicator function. To maintain the portfolio context we can interpret (17) as a process for the cross-sectionally aggregated returns of the assets in the portfolio.[1] Deriving higher moments of temporally aggregated multivariate GARCH models is technically demanding and to a large extent an unexplored field, though, and we view it as beyond the scope of this particular study (see Hafner, 2003, 2008, for some results).

The conditional variance specification in (17) is the asymmetric GARCH model of Glosten et al. (1993). The term, $\gamma\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$, in (17) extends the basic GARCH$(1,1)$ of Bollerslev (1986) and captures the leverage effect in financial markets, i.e. the asymmetric response of future volatility to positive and negative shocks. This feature has empirically been found highly relevant and several other models to cope with it exist. Wong and So (2003) consider for example the QGARCH model of Sentana (1995) and Engle (1990). The most popular model in empirical work appears, however, to be the GJR-GARCH. In fact, among several different asymmetric GARCH models applied to Japanese stock index data Engle and Ng (1993) found that the best performing parametric specification indeed was the GJR-GARCH.

The implied $VaR_{T,k}^{1-\alpha}$ and $ES_{T,k}^{1-\alpha}$ are given by

$$VaR_{T,k}^{1-\alpha} = -q_{T,k}^{\alpha}\sqrt{h_{T,k}} \tag{18}$$

$$ES_{T,k}^{1-\alpha} = -e_{T,k}^{\alpha}\sqrt{h_{T,k}}. \tag{19}$$

Direct calculation give that the multiple period conditional variance of $Y_{T,k}$ is given by

$$h_{T,k} = \frac{\omega}{1-(\alpha+\beta+\gamma/2)}\left(k - \frac{1-(\alpha+\beta+\gamma/2)^k}{1-(\alpha+\beta+\gamma/2)}\right) + \frac{1-(\alpha+\beta+\gamma/2)^k}{1-(\alpha+\beta+\gamma/2)}h_{T+1}.$$

[1] Berkowitz and O'Brien (2002) used a similar approach to study the accuracy of the VaR's reported by commercial banks.

The analytical approximations to $q^{\alpha}_{T,k}$ and $e^{\alpha}_{T,k}$ in (18) and (19) require that we compute theoretical conditional moments of $Y_{T,k}$. We restrict ourselves to the third and the fourth conditional moments and in the Appendix we show how these may be obtained. The corresponding conditional moments of $\varepsilon_{T,k}$ are then

$$\begin{aligned} s_{T,k} &= \frac{E_T(Y^3_{T,k})}{h^{3/2}_{T,k}}, \\ k_{T,k} &= \frac{E_T(Y^4_{T,k})}{h^2_{T,k}}. \end{aligned}$$

When $\gamma = 0$, the model (17) simplifies to the basic GARCH$(1,1)$ model. Breuer and Jandačka (2007) give expressions for the conditional variance, $h_{T,k}$, and the conditional kurtosis, $k_{T,k}$, of $Y_{T,k}$ under GARCH$(1,1)$ variance.

When $\gamma \neq 0$, there is conditional skewness in $Y_{T,k}$ and the derivation involves non-integer moments $E_T\left(h^{3/2}_{T+k}\right)$ and $E_T\left(h^{5/2}_{T+k}\right)$. Non-integer moments also arise in the context of option pricing in the GARCH framework in Duan, Gauthier, Simonato, and Sasseville (2006), who use Taylor expansions to approximate $E_T\left(h^{1/2}_{T+k}\right)$ and $E_T\left(h^{3/2}_{T+k}\right)$. This is the route taken here as well and the natural starting point for the expansions in our conditional setting is the conditional expectation of the future conditional variance, i.e. $E_T\left(h_{T+k}\right)$. The approximations would then have the form $E_T\left(h^i_{T+k}\right) = a_1 + a_2 E_T\left(h^2_{T+k}\right) + \ldots$, where $i = 3/2, 5/2$ and the a's are functions of $E_T\left(h_{T+k}\right)$. An important issue is whether higher integer moments of h_{T+k} exist or not for a particular process. Ling and McAleer (2002) derive necessary and sufficient conditions for the unconditional expectation of h^m_{T+k}, m integer, to exist for the family of GARCH$(1,1)$ processes in He and Teräsvirta (1999). The family nests the GJR-GARCH and if $E\left(|\varepsilon|^{2m}\right) < \infty$ the conditions for that particular model are $\omega^m < \infty$ and $E\left[\left(\beta + [\alpha + \gamma \mathbf{1}\left(\varepsilon_{t-1} < 0\right)]\varepsilon^2_{t-1}\right)\right]^m < 1$. The condition for the unconditional variance of y_t to exist is for example $\omega/(1 - \alpha - \beta - \gamma/2) > 0$. Even though the setting here is conditional these conditions can potentially put restrictions on the applicability of our approximation approaches as they require computation of higher moments of y_t. Here, we consider second order expansions.

The approximations based on the Gram-Charlier expansion and the Root-k need no additional comments and are directly obtained by plugging in the expressions for $s_{T,k}$ and $k_{T,k}$. When it comes to choosing a distribution for $\varepsilon_{T,k}$ in the second approach our only requirement is that the first five moments exist for the distribution, and we thus have a large menu to choose from. In the finance literature several distributions have been studied in the context of allowing for conditional skewness and excess kurtosis. Harvey and Siddique (1999) consider a non-central t distribution. Brännäs and Nordman (2003) study the Pearson type IV and the log-generalized gamma. Given that the requirement is satisfied it is difficult to ex ante argue in favor of one distribution over another. A distribution that has gained increasing popularity in the literature (e.g. Jondeau and Rockinger, 2006) is the skewed Student's t distribution of Hansen (1994). Wong and So (2003) propose the distribution in Theodossiou (1998), which is similar to the one in Hansen (1994). They do not pursue the analysis allowing for skewness though and restrict themselves to the symmetric Student's t distribution.

The *pdf* of a zero mean and unit variance skew-t distributed variable, Z, is

$$g(z) = \begin{cases} bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2}, & z < -a/b, \\ bc\left(1 + \frac{1}{\nu-2}\left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2}, & z \geq -a/b, \end{cases} \tag{20}$$

where $2 < \nu < \infty$, $-1 < \lambda < 1$, $a = 4\lambda c(\nu-2)/(\nu-1)$, $b^2 = 1 + 3\lambda^2 - a^2$ and $c = \Gamma[(\nu+1)/2]/\left[\sqrt{\pi(\nu-2)}\Gamma(\nu/2)\right]$. In this particular case the approach consists of matching $s_{T,k}$ and $k_{T,k}$ to the corresponding moments of the skew-t distribution. Jondeau and Rockinger (2003) show that the third and fourth moments of the skew-t distribution are given by

$$\begin{aligned} E\left(Z^3\right) &= \left(m_3 - 3a\,m_2 + 2a^3\right)/b^3, \\ E(Z^4) &= \left(m_4 - 4a\,m_3 + 6a^2 m_2 - 3a^4\right)/b^4, \end{aligned}$$

where $m_2 = 1 + 3\lambda^2$, $m_3 = 16c\,\lambda(1+\lambda^2)(\nu-2)^2/[(\nu-1)(\nu-3)]$ and $m_4 = 3(\nu-2)(1 + 10\lambda^2 + 5\lambda^4)/(\nu-4)$. The third moment is defined for $\nu > 3$, while the fourth is defined for $\nu > 4$. The implied values on λ and ν are then obtained as the solution in terms of $s_{T,k}$ and $k_{T,k}$ to

$$\begin{aligned} s_{T,k} &= \left(m_3 - 3a\,m_2 + 2a^3\right)/b^3 \\ k_{T,k} &= \left(m_4 - 4a\,m_3 + 6a^2 m_2 - 3a^4\right)/b^4. \end{aligned} \tag{21}$$

Except for the symmetric case, i.e. $\lambda = 0$, (when $\nu = (6 - k_{T,k})/(3 - k_{T,k})$) we were not able to derive λ and ν as nice explicit functions of $s_{T,k}$ and $k_{T,k}$. Obtaining the values then amounts to solving the system numerically.[2] Of course, the valid region for λ and ν also implies a region in the $s_{T,k}$ and $k_{T,k}$ dimension. Jondeau and Rockinger (2003) note that the relation between these regions is bijective when $\nu > 4$ and $|\lambda| < 1$ implying that the solution to (21) is unique.

To compute the VaR and ES we require expressions for $\tilde{q}^{\alpha}_{T,k}$ and $\tilde{e}^{\alpha}_{T,k}$ as inputs to (9) and (10), respectively. Jondeau and Rockinger (2003) show that the αth quantile of the skew-t distribution is given by

$$q_\alpha = \begin{cases} \frac{1}{b}\left[(1-\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha}{1-\lambda}\right) - a\right] & \text{if } \alpha < \frac{1-\lambda}{2}, \\ \frac{1}{b}\left[(1+\lambda)\sqrt{\frac{\nu-2}{\nu}}F^{-1}\left(\frac{\alpha+\lambda}{1+\lambda}\right) - a\right] & \text{if } \alpha \geq \frac{1-\lambda}{2}. \end{cases}.$$

[2] In the simulation study in Section 5 we employed the following solver

$$\begin{aligned} (\lambda,\nu)' = & \underset{-1<\lambda<1,\nu>4}{\arg\min} \left[s_{T,k} - \left(m_3 - 3a\,m_2 + 2a^3\right)/b^3\right]^2 \\ & + \left[k_{T,k} - \left(m_4 - 4a\,m_3 + 6a^2 m_2 - 3a^4\right)/b^4\right]^2 \end{aligned}$$

Issues with using a solver of this type are discussed in Press, Teukolsky, Vetterling, and Flannery (2007, ch. 9). However, it performed satisfactory in our application with function values close to zero. We also compared it to the Newton-Raphson algorithm in Press et al. (2007, ch. 9, p. 475) and almost identical values were obtained. The latter was highly sensitive to the starting values, though.

In the Appendix we show that, for $\alpha < (1-\lambda)/2$

$$\begin{aligned} E\left(\varepsilon \mid \varepsilon \leq q_\alpha\right) &= -\frac{(1-\lambda)^2}{\alpha b}\left[\frac{\sqrt{v(\nu-2)}}{\nu-1}\left(1+\frac{q_\alpha^{*2}}{\nu}\right) f\left(q_\alpha^*\right)\right. \\ &\quad \left. +\frac{a}{1-\lambda} F\left(q_\alpha^*\right)\right], \end{aligned}$$

where $q_\alpha^* = (bq+a)\sqrt{\nu/(\nu-2)}/(1-\lambda)$ and f and F are the pdf and cdf of the Student's t distribution.

Quantifying the uncertainty of the predictors follows from Section 3.

5 Simulation study

The discussion regarding the approximative predictors has so far been theoretical, but what is of obvious practical interest is their properties in finite sample. We address this question by means of quite detailed Monte Carlo simulations based on the model in (17). The study was carried out using the RATS 6.30 package. To estimate the GJR-GARCH models we employed the built-in GARCH procedure with the BFGS-algorithm, but as the variance-covariance estimator we used $T^{-2}(\partial \ln L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}\ \partial \ln L_T(\hat{\boldsymbol{\theta}})/\partial\boldsymbol{\theta}')$.

When it comes to designing the experiment we note for the variance specification that the degree of persistence and asymmetry are of particular interest. In a related study Christoffersen and Gonçalves (2005) simulate the GARCH$(1,1)$-model with $\omega = (1-\alpha-\beta)20^2/252$, $\alpha = 0.1$ and persistence parameter $\beta = 0.4, 0.8$ and 0.89. Here, the additional parameter γ introduces asymmetry and we consider three degrees: $(\alpha,\gamma) = (0.1, 0)$, $(0.05, 0.1)$ and $(0, 0.2)$. The unconditional variance is thus the same throughout. For estimation we use samples of sizes 500 and 1000, which are realistic sample sizes corresponding to approximately 2 and 4 years of daily trading data. For the simulation based predictor we use $R = 100\ 000$ to isolate the effect of the estimation error in $\hat{\boldsymbol{\theta}}$. The results are based on $N = 1000$ replications. Note however, that we discard without replacement the cases when the ML estimator did not converge to a valid point or when an approximation failed for some reason. Table A1 in the Appendix gives the proportions of cases when this happened. The remaining design parameters are the confidence level and the horizon. Increasing the confidence level means that we make predictions further out in the left tail, which intuitively increases the uncertainty. Predicting further into the future is also associated with greater uncertainty, which should be reflected in the performance of the predictors. We set the confidence level to either 95% or 99% and consider $k = 5$ and 10. In Table 1 we give bias, mean square errors (MSE) and estimated asymptotic variances (EAV) for the case $\beta = 0.8$ and $(\alpha,\gamma) = (0.05, 0.1)$. The tables for the other parameter combinations are given in the Appendix.

We make no distinction between VaR and ES in the discussion as the results are qualitatively similar. Considering first the bias we see that it is largest and negative for the Root-k approach. The bias for the G-C approach is positive for all cases and surprisingly large for the higher confidence level. Overall, it is the smallest for the W-S and the simulation based approaches. With some exceptions, the bias gets more pronounced when increasing the confidence level and the horizon, and it decreases when increasing the sample size. Turning to the accuracy in terms of MSE we again have a rather clear ranking with the Root-k approach being the worst and the W-S and the simulation based approach tied in first place. Without exceptions, the qualitative effects of the design variables are the same as for the bias. Of interest for the computation of, e.g., prediction intervals is how well the delta method approximates the finite sample variance of the predictors. To scrutinize on this we may compare the MSE to the average of the corresponding estimated asymptotic variances.[3] The delta method appears to perform quite satisfactorily for the GC, the W-S and the simulation based approaches. Regarding the Root-k approach it is difficult to draw any conclusions due to the often large bias.

In a smaller scale experiment we examined the robustness of the results for data generated according to a GJR-GARCH process ($\alpha = .05$, $\beta = 0.8$ and $\gamma = 0.1$) and with skew–t innovations ($\lambda = -0.2$ and $\nu = 8$). We computed the predictors for a confidence level of 95% and $k = 5$ based on $T = 1000$ both with a correct distributional assumption and with an incorrect assumption of normality (cf. QMLE). Regarding the former the predictors need to be adapted to the skew-t distribution and the corresponding derivations may be found in the Appendix.

The results are given in Table A10 of the Appendix and, with some exceptions, they are qualitatively similar to the ones above when the model is correctly specified. However, it is noteworthy that the delta method appears to work poorly in many cases. Also, the bias and the MSE for the GC-approach in case of ES prediction is very high. Under an incorrect normality assumption the bias is negative in all cases and the ranking is different.

An important question we wish to answer is that of which method is the best. For this we use a Diebold-Mariano type of test (Diebold and Mariano, 1995). They show that the predictive superiority of one predictor over another can be tested by means of a simple t-test of the standardized difference between the loss functions. Here, the loss function is the squared prediction error and the test statistic was computed as the t-statistic in the regression of the pooled differences on a constant. To take care of heteroskedasticity we used Eicker-White standard errors. In Table 2 we give results for all pairwise tests.

Among the analytical approaches the one based on the skew-t distribution is judged the best. In fact, it also fares better than the simulation based in case of ES prediction. For VaR, all analytical approaches are rejected in favor of the simulation based. The actual differences between the simulation based and the one based on the skew-t is small, though. In a practical situation one would thus supposedly prefer the latter thanks to its advantage in computing time. For example, consider the task of computing VaR and ES for confidence levels 95% and 99% and horizons of 5 and 10 periods given that parameter

[3]Here we rely on a central limit theorem argument. Let $\{Z_n\}_{n=1}^{N}$ be independently distributed variables with zero means and variances σ_n^2. Then $\sum_{n=1}^{N} Z_n^2/N$ is a consistent estimator of $\bar{\sigma}^2 = \sum_{n=1}^{N} \sigma_n^2/N$. Hence, when the bias is small the average of the estimated asymptotic variances should be close to the MSE.

Table 1: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.05y_{t-1}^2 + 0.8h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance. The averages are for the true values.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Average	4.7041			6.7147			6.2311			8.9773		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.1109	0.1524	0.1243	-0.2200	0.4417	0.2486	-0.4707	0.4247	0.1955	-0.8322	1.2222	0.3910
G-C	0.0357	0.1401	0.1404	0.0675	0.3247	0.3081	0.2340	0.5080	0.5319	0.5437	1.8331	1.8692
S-W	0.0280	0.1364	0.1329	0.0446	0.3062	0.2810	-0.0058	0.2776	0.2818	-0.0068	0.6915	0.6743
Sim	-0.0115	0.1301	0.1484	-0.0203	0.2876	0.3447	-0.0195	0.2753	0.2869	-0.0429	0.6780	0.6704
						$T=1000$						
Root-k	-0.0995	0.0734	0.0572	-0.2028	0.2723	0.1145	-0.4567	0.2931	0.0900	-0.8112	0.9393	0.1801
G-C	0.0456	0.0605	0.0652	0.0873	0.1454	0.1446	0.2231	0.2403	0.2238	0.5176	0.9450	0.7638
S-W	0.0387	0.0586	0.0621	0.0669	0.1348	0.1326	0.0091	0.1199	0.1304	0.0243	0.3057	0.3154
Sim	-0.0010	0.0545	0.0673	-0.0007	0.1238	0.1577	-0.0013	0.1179	0.1339	-0.0063	0.2994	0.3164
	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Average	7.1562			10.3360			8.6290			12.5751		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.6596	0.6892	0.2486	-1.1496	1.9641	0.4973	-1.1858	1.7309	0.3264	-2.0496	4.9457	0.6527
G-C	0.1247	0.5109	0.5478	0.2657	1.5148	1.5921	0.3088	0.9728	0.9175	0.3404	2.1102	2.1023
S-W	-0.0211	0.3923	0.3982	-0.0259	0.9982	0.9835	-0.0896	0.6902	0.7208	-0.1418	1.9228	1.9354
Sim	-0.0231	0.3947	0.4470	-0.0555	0.9890	1.1618	-0.0303	0.7290	0.7875	-0.0808	1.9926	2.0277
						$T=1000$						
Root-k	-0.6438	0.5185	0.1145	-1.1261	1.5978	0.2290	-1.1681	1.4975	0.1503	-2.0237	4.4416	0.3006
G-C	0.1358	0.2279	0.2434	0.2829	0.6916	0.6898	0.3652	0.5425	0.4611	0.5443	1.2579	1.0328
S-W	-0.0031	0.1706	0.1839	0.0116	0.4435	0.4598	-0.0688	0.3047	0.3285	-0.0998	0.8602	0.9024
Sim	-0.0008	0.1700	0.2055	-0.0084	0.4368	0.5428	0.0010	0.3173	0.3656	-0.0147	0.8926	0.9634

Table 2: Diebold-Mariano t-tests. Positive values are in favor of the method in the second row. The loss function is the squared prediction error and the statistics were computed from the regression of the pooled differences on a constant using Eicker-White standard errors. The average differences are given in parentheses.

	Root-k	Root-k	Root-k	G-C	G-C	S-W
vs.	G-C	S-W	Sim	S-W	Sim	Sim
VaR	48.441	72.704	74.094	30.599	31.051	8.980
	(0.351)	(0.467)	(0.471)	(0.117)	(0.121)	(0.005)
ES	50.354	107.547	107.019	44.180	43.808	-4.231
	(0.853)	(1.396)	(1.390)	(0.542)	(0.536)	(-0.004)

Table 3: Effects of design variables on the accuracy of the predictors. The numbers are the values on t-tests of zero coefficents in dummy variable regressions, where the base case is $T = 500$, $\alpha = 0.05$, $k = 5$, $\beta = 0.8$ and $\gamma = 0.1$.

	Root-k		G-C		S-W		Sim	
Dummy	VaR	ES	VaR	ES	VaR	ES	VaR	ES
$T = 1000$	-12.438	-6.576	-33.782	-24.312	-37.497	-41.051	-39.098	-40.875
$\alpha = .01$	147.422	148.049	79.504	34.673	64.808	56.867	66.674	59.286
$k = 10$	79.897	90.368	52.018	54.081	47.605	50.566	47.553	50.450
$\beta = .4$	-17.208	-47.475	-28.518	-31.511	-22.629	-22.056	-24.940	-22.676
$\beta = .89$	-19.582	-11.402	-16.423	-12.319	-16.621	-19.465	-15.831	-19.242
$\gamma = .0$	-55.443	-88.258	-23.289	-38.378	-19.264	-26.797	-16.953	-25.971
$\gamma = .2$	45.713	67.273	11.665	25.276	4.441	7.945	2.382	5.776

estimates have been obtained. Along with standard errors it takes approximately 25 seconds on a 1.83 GHz Intel Centrino Duo processor employing the simulation based approach, while the other approaches compute the quantities within the blink of an eye.

Of further practical interest is how the prediction accuracy varies with the design variables, i.e. the sample size, confidence level, horizon and model parameters. For this we ran the dummy variable regressions

$$\begin{aligned} \ln(\widehat{VaR}_{T,k}^{1-\alpha} - VaR_{T,k}^{1-\alpha})^2 &= \boldsymbol{\beta}_v'\mathbf{d}+\xi_v, \\ \ln(\widehat{ES}_{T,k}^{1-\alpha} - ES_{T,k}^{1-\alpha})^2 &= \boldsymbol{\beta}_e'\mathbf{d}+\xi_e, \end{aligned}$$

where $\mathbf{d}$ is a vector of dummy variables indicating value on design variable and ξ_v and ξ_e are the error terms. We again used Eicker-White standard erros. The base case is taken to be $T = 500$, $\alpha = .05$, $k = 5$, $\beta = .8$ and $\gamma = .1$. We ran one regression for each method and the results are given in Table 3.

The results were uniform across the methods and qualitatively the same for VaR and ES. Not surprisingly, doubling the sample size significantly increased the accuracy.

The predictions at confidence level 99% were significantly less accurate than the ones at the 95% level. Increasing the horizon from 5 to 10 periods significantly decreased the accuracy. Both when increasing and reducing the persistence in the conditional variance the accuracy is significantly enhanced compared to the base case. Regarding the effects of asymmetry we note that predictions for the no asymmetry case (i.e. standard GARCH) are significantly more accurate than the ones for the base case. The opposite is true for the case when only negative shocks affects the future variance.

6 Empirical illustration

In this section we provide a small illustration of the above approximation approaches, where the object of interest is the five day VaR of the S&P 500 index. Eight years of daily data were downloaded from DataStream and the sample covers October 31, 2000 to October 31, 2008, for a total of 2089 daily observations on the index.

Returns were calculated as $y_t = 100 \times \log(I_t/I_{t-1})$, where I_t is the value of the index at t. We assume that daily returns are generated by a GJR-GARCH process with a constant mean and standard normally distributed shocks. In the estimation of the model as is we often obtained a negative coefficient on the squared residual. This causes problems for the simulation based predictor, since the conditional variance may become negative in the out of sample simulations. To force positive variances we adopted the following version

$$h_t = \omega + \exp(\alpha) u_{t-1}^2 + \beta h_{t-1} + \gamma \mathbf{1}\left(u_{t-1} < 0\right) u_{t-1}^2,$$

where u_t is the one-period ahead prediction error. Regarding the computation of VaR a comment is in place. Recall the decomposition $Y_{T,k} = \mu_{T,k} + \varepsilon_{T,k}\sqrt{h_{T,k}}$. As inputs to the G-C and W-S approximations we require the conditional skewness and kurtosis of $\varepsilon_{T,k}$. Those were derived in the Appendix under a zero conditional mean of $Y_{T,k}$. Here, we use the same derivation but replace $Y_{T,k}$ with $Y_{T,k} - \mu_{T,k}$, where $\mu_{T,k} = k\mu$. Note also that the uncertainty in μ should be recognized in the computation of the variances of VaR.

Based on a rolling prediction scheme we obtained VaR predictions at the confidence level 95% and in estimation we considered samples of size 500 observations. We discarded cases when the computation of the predictors failed for some reason and obtained 1522 predictions. Robust standard errors of the sandwich form were employed throughout. The final successfully estimated model (October 3, 2008) on the implied conventional form is given below along with some diagnostics.

$$
\begin{aligned}
y_t &= \underset{(-0.52)}{-0.018} + u_t, \\
h_t &= \underset{(1.95)}{0.014} + \underset{(0.10)}{1 \times 10^{-5}} u_{t-1}^2 + \underset{(36.88)}{0.891} h_{t-1} + \underset{(4.99)}{0.211} u_{t-1}^2 \mathbf{1}(u_{t-1} < 0), \\
L &= -662.24,\ LB_{10} = 8.22,\ LB_{10}^2 = 6.95,\ JB = 0.29,
\end{aligned}
$$

where t-statistics are given in parentheses, L is the value of the log-likelihood function, LB_{10} and LB_{10}^2 give the values of the test-statistics in the Ljung-Box test of no auto-correlation up to lag 10 in standardized residuals and squared standardized residuals, respectively, and JB is the value of the test-statistic in the Jarque-Bera normality-test. The conditional variance is highly persistent and the asymmetric effect of past shocks is considerable. Noteworthy is also that there is no remaining ARCH-effect in the standardized residuals and that normality is not rejected.

When it comes to assessing the performance of the VaR predictors we follow the likelihood ratio framework of Christoffersen (1998). Let P denote the number of VaR predictions and let H_t, $t = 1, ..., P$, denote the hit sequence, i.e. $H_t = 1$ if the actual return exceeds the predicted VaR and is 0 otherwise. For a good VaR predictor the unconditional exceedence rate, $\hat{\alpha} = \sum H_t/P$, should be close to α. This can be tested by the statistic $LR_{unc} = -2\ln[(1-\alpha)^{P-H}\alpha^H] + 2\ln[(1-\hat{\alpha})^{P-H}\hat{\alpha}^H]$. Christoffersen (1998) notes that the hit sequence should not only sum up to αP, but also be an iid Bernoulli sequence with parameter α. As a test of independence he proposes the test statistic $LR_{ind} = -2\ln[(1-\hat{\alpha})^{P_{00}+P_{10}}\hat{\alpha}^{P_{01}+P_{11}}] + 2\ln[(1-\hat{\pi}_0)^{P_{00}}\hat{\pi}_0^{P_{01}}(1-\hat{\pi}_1)^{P_{10}}\hat{\pi}_1^{P_{11}}]$, where P_{ij} is defined as the number of periods in which state j occurred in one period, while state i occurred the previous period and π_i is the probability of a hit conditional on state i the previous day.[4] He proposes $LR_{cc} = LR_{unc} + LR_{ind}$ as a statistic for the joint test of correct conditional coverage. Asymptotically LR_{unc} and LR_{ind} are $\chi^2(1)$-distributed, while LR_{cc} is $\chi^2(2)$-distributed. Our multiple period context may give rise to serial dependence in the raw hit sequence. To cope with the problem we use Bonferroni subsamples (Dowd, 2007). Thus, the raw hit sequence is split up into five hit sequences and the statistics are computed for each sequence. We reject an overall test at significance level λ if the test is rejected for any of the subsamples using level $\lambda/5$.

The unconditional exceedence rates for the predictors are 5.532%, 5.506%, 5.506% and 5.512% for the Root-k, G-C, W-S and simulated based approach, respectively. In Figure 1 we display VaR's and standard errors for the turbulent period September 5, 2008 to October 3, 2008.

To digress further on the performance of the predictors we present in Table 4 the results of the backtesting of the VaR predictors. The approaches perform very similarly and no results are significant at conventional levels. Note that the performances of all predictors are quite weak when the prediction origin is a Monday (too many hits) or a Thursday (too few hits).

[4] When $\hat{\pi}_1 = 0$ we used $LR_{ind} = -2\ln[(1-\hat{\alpha})^{P_{00}+P_{10}}\hat{\alpha}^{P01+P11}] + 2\ln[(1-\hat{\pi}_0)^{P_{00}}\hat{\pi}_0^{P_{01}}$ (cf. Christoffersen and Pelletier, 2004).

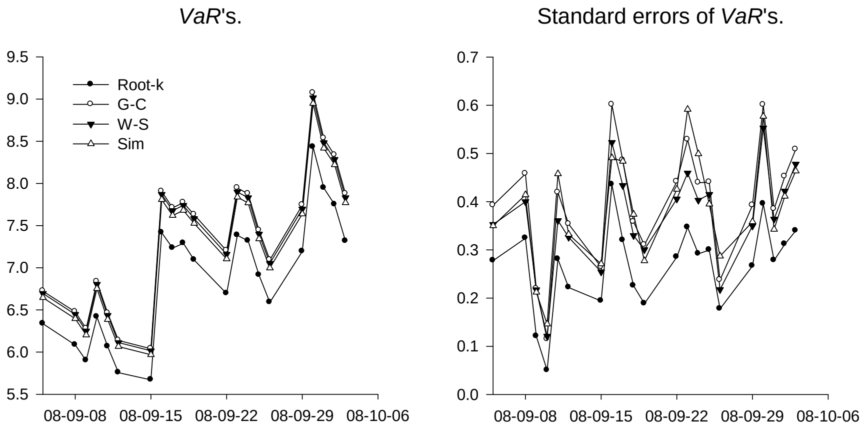

Figure 1: VaR's for 100× log-returns for the S&P 500 index.

Table 4: Backtesting of the VaR predictors. The top row indicate day of the week of the prediction origin.

	Root-k					G-C				
	M	T	W	T	F	M	T	W	T	F
$\hat{\alpha}$	0.0689	0.0498	0.0525	0.0391	0.0559	0.0656	0.0465	0.0492	0.0391	0.0526
LR_{unc}	2.0525	0.0002	0.0382	0.8293	0.2165	1.4244	0.0789	0.0043	0.8293	0.0436
LR_{ind}	0.0540	0.1275	0.0640	0.8782	1.6185	0.0311	1.2665	1.4333	0.8782	1.4223
LR_{cc}	2.1064	0.1277	0.1022	1.7075	1.8350	1.4555	1.3453	1.4376	1.7075	1.4659
	S-W					Sim				
	M	T	W	T	F	M	T	W	T	F
$\hat{\alpha}$	0.0656	0.0465	0.0492	0.0391	0.0526	0.0656	0.0465	0.0492	0.0391	0.0559
LR_{unc}	1.4244	0.0789	0.0043	0.8293	0.0436	1.4244	0.0789	0.0043	0.8293	0.2165
LR_{ind}	0.0311	1.2665	1.4333	0.8782	1.4223	0.0311	1.2665	1.4333	0.8782	1.6185
LR_{cc}	1.4555	1.3453	1.4376	1.7075	1.4659	1.4555	1.3453	1.4376	1.7075	1.8350

7 Conclusions

In this paper we studied four methods to approximate VaR and ES for multiple period returns. We also viewed the uncertainty arising from the estimation error important and we discussed how to employ the delta method to quantify this uncertainty. Based on the result of a simulation experiment we conclude that among the approaches studied the one based on assuming a skew-t distribution for the multiple period returns and that based on simulations were the best. The predictors based on the Root-k and the Gram-Charlier showed positive and negative bias, respectively. Except for the Root-k approach we found that the uncertainty due to the estimation error can be quite accurately estimated employing the delta method.

In an empirical illustration we computed 5 day VaR's for the S&P 500 index using the approximative predictors. In terms of exceedence rates all approaches performed similarly and we could not reject any of them at conventional significance levels.

Appendix

Conditional moments of $Y_{T,k}$ with GJR-GARCH conditional variance

We consider here the case $E_T(y_{T+i}) = 0$ and when deriving the conditional moments it is helpful to use the decomposition $Y_{T,k} = \sum_{i=1}^{k-1} y_{t+i} + y_{t+k}$. Let $s = E(\varepsilon^3_{T+i})$ and $\kappa = E(\varepsilon^4_{T+i})$. For notational convenience we let $\varepsilon = \varepsilon_{T+i}$. Obtaining the moments then amounts to solving the system

$$E_T\left[\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2\right] = E_T\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2 + E_T\left(h_{T+k-1}\right) \tag{A1}$$

$$\begin{aligned} E_T\left[\left(\sum_{i=1}^{k} y_{T+i}\right)^3\right] &= sE_T(h^{3/2}_{T+k}) + E_T\left(\sum_{i=1}^{k-1} y_{T+i}\right)^3 \\ &\quad +3E_T\left(h_{T+k}\sum_{i=1}^{k-1} y_{T+i}\right) \end{aligned} \tag{A2}$$

$$\begin{aligned} E_T\left[\left(\sum_{i=1}^{k} y_{T+i}\right)^4\right] &= \kappa E_T h^2_{T+k} + E_T\left(\sum_{i=1}^{k-1} y_{T+i}\right)^4 \\ &\quad +4sE_T\left(h^{3/2}_{T+k}\sum_{i=1}^{k-1} y_{T+i}\right) \\ &\quad +6E_T h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2 \end{aligned} \tag{A3}$$

$$E_T\left(h_{T+k}\right) = \omega + \delta E_T h_{T+k-1} \tag{A4}$$

$$\begin{aligned} E_T\left(h^{3/2}_{T+k}\right) &\approx \frac{5}{8}(E_T h_{T+k})^{3/2} \\ &\quad +\frac{3}{8\sqrt{E_T h_{T+k}}} E_T h^2_{T+k} \end{aligned} \tag{A5}$$

$$\begin{aligned} E_T\left(h^2_{T+k}\right) &= \omega^2 + 2\omega\delta E_T\left(h_{T+k-1}\right) \\ &\quad +\lambda E_T h^2_{T+k-1} \end{aligned} \tag{A6}$$

$$\begin{aligned} E_T\left(h^{5/2}_{T+k}\right) &\approx -\frac{7}{8}(E_T h_{T+k})^{5/2} \\ &\quad +\frac{15}{8}\sqrt{E_T h_{T+k}} E_T h^2_{T+k} \end{aligned} \tag{A7}$$

$$\begin{aligned} E_T\left(h_{t+k}\sum_{i=1}^{k-1} y_{t+i}\right) &= [\alpha s + \gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^3)] E_T h^{3/2}_{T+k-1} \\ &\quad +\delta E_T\left(h_{k+t-1}\sum_{i=1}^{k-2} y_{i+t}\right) \end{aligned} \tag{A8}$$

$$E_T\left[h^{3/2}_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)\right] \approx \frac{3}{4}(E_T h_{T+k})^{1/2} E_T\left[h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)\right]$$

$$
\begin{aligned}
& +\frac{3}{8\sqrt{E_T h_{T+k}}} E_T\left(h_{T+k}^2 \sum_{i=1}^{k-1} y_{T+i}\right) & \text{(A9)} \\
E_T\left(h_{T+k}^2 \sum_{i=1}^{k-1} y_{T+i}\right) = \; & 2\omega\delta E_T\left(h_{T+k-1} \sum_{i=1}^{k-2} y_{T+i}\right) \\
& +\lambda E_T\left(h_{T+k-1}^2 \sum_{i=1}^{k-2} y_{T+i}\right) \\
& +\pi E_T\left(h_{T+k-1}^{5/2}\right) \\
& +[2\alpha\omega s + 2\gamma\omega E(\mathbf{1}(\varepsilon < 0)\varepsilon^3)] \\
& \times E_T\left(h_{T+k-1}^{3/2}\right) & \text{(A10)} \\
E_T\left[h_{T+k}\left(\sum_{i=1}^{k-1} y_{T+i}\right)^2\right] = \; & \omega E_T\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2 \\
& +\delta E_T\left[h_{T+k-1}\left(\sum_{i=1}^{k-2} y_{T+i}\right)^2\right] \\
& +[2\alpha s + 2\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^3)] \\
& \times E_T\left[h_{T+k-1}^{3/2}\left(\sum_{i=1}^{k-2} y_{T+i}\right)\right] \\
& +\omega E_T h_{T+k-1} + \mu E_T\left(h_{T+k-1}^2\right), & \text{(A11)}
\end{aligned}
$$

where $\mathbf{1}(\cdot)$ is the indicator function, $\delta = \alpha + \beta + \gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^2)$, $\lambda = \beta^2 + \kappa\alpha^2 + 2\alpha\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^4) + 2\alpha\beta + 2\beta\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^2) + \gamma^2 E(\mathbf{1}(\varepsilon < 0)\varepsilon^4)$, $\mu = \kappa\alpha + \beta + \gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^4)$ and $\pi = \alpha^2 E(\varepsilon^5) + 2\alpha\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^5) + 2\alpha\beta s + 2\beta\gamma E(\mathbf{1}(\varepsilon < 0)\varepsilon^3) + \gamma^2 E(\mathbf{1}(\varepsilon < 0)\varepsilon^5)$.

When ε is Gaussian we have $s = E(\varepsilon^5) = 0$ and $E(\varepsilon^4) = 3$. Also, it is straightforward to show that $E(\mathbf{1}(\varepsilon < 0)\varepsilon) = -\phi(0)$. For integer $r > 1$ it holds that $\int_{-\infty}^0 z^r \phi(z)dz = (r-1)\int_{-\infty}^0 z^{r-2}\phi(z)dz$. We have $E(\mathbf{1}(\varepsilon < 0)\varepsilon^2) = 1/2$, $E(\mathbf{1}(\varepsilon < 0)\varepsilon^3) = -2\phi(0)$, $E(\mathbf{1}(\varepsilon < 0)\varepsilon^4) = 3/2$ and $E(\mathbf{1}(\varepsilon < 0)\varepsilon^5) = -8\phi(0)$.

Properties of the skew-t distribution

We take a, b, c, m_2, m_3 and m_4 as they are given in the text. Jondeau and Rockinger (2003) give the αth quantile of the skew-t distribution as

$$
q_\alpha = \begin{cases} \frac{1}{b}\left[(1-\lambda)\sqrt{\frac{\nu-2}{\nu}} F^{-1}\left(\frac{\alpha}{1-\lambda}\right) - a\right] & \text{if } \alpha < \frac{1-\lambda}{2}, \\ \frac{1}{b}\left[(1+\lambda)\sqrt{\frac{\nu-2}{\nu}} F^{-1}\left(\frac{\alpha+\lambda}{1+\lambda}\right) - a\right] & \text{if } \alpha \geq \frac{1-\lambda}{2}, \end{cases}
$$

where $F^{-1}(\cdot)$ is the inverse of the cdf of the Student's t distribution with ν degrees of freedom.

To solve the system (A1) - (A11) we require some integer moments. We first derive the censored ones for the standardized Student's t distribution. Let $\mu_m^q =$

$\int_{-\infty}^{q} x^m t\left(x\right) dx$, where $t(x) = c\left[1 + x^2/(\nu-2)\right]^{-(\nu+1)/2}$. μ_0^q is obvious and adapting a result in Andreev and Kanto (2005) gives

$$\mu_1^q = -\frac{\nu-2}{\nu-1}\left(1+\frac{q^2}{\nu-2}\right) t\left(q\right).$$

By integration by parts we have for $m > 1$

$$\begin{aligned}
\mu_m^q &= \int_{-\infty}^{q} x^m c\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2} dx \\
&= \left\{-x^{m-1}\frac{\nu-2}{\nu-1}\left[\left(1+\frac{x^2}{\nu-2}\right)t(x)\right]\right\}_{-\infty}^{q} \\
&\quad +(m-1)\frac{\nu-2}{\nu-1}\int_{-\infty}^{q} x^{m-2}\, c\left(1+\frac{x^2}{\nu-2}\right)t(x)dx \\
&= q^{m-1}\mu_1^q + (m-1)\frac{\nu-2}{\nu-1}\left(\mu_{m-2}^q + \frac{\mu_m^q}{\nu-2}\right).
\end{aligned}$$

Then

$$\mu_m^q = \frac{\nu-1}{\nu-m}\left(q^{m-1}\mu_1^q + (m-1)\frac{\nu-2}{\nu-1}\mu_{m-2}^q\right).$$

Now, for the skew-t distributed variable Z we have for $q < -a/b$

$$\begin{aligned}
E(\mathbf{1}(Z \le q)Z^m) &= \int_{-\infty}^{q} z^m bc\left(1+\frac{1}{\nu-2}\left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} dz \\
&= \frac{\lambda_*}{b^m}\int_{-\infty}^{q_*}\left[\lambda_* y - a)\right]^m t(y)dy \\
&= \frac{\lambda_*}{b^m}\int_{-\infty}^{q_*}\sum_{i=0}^{m}\left(\begin{array}{c} m \\ i \end{array}\right)\lambda_*^{m-i}(-a)^i y^{m-i} t(y)dy \\
&= \frac{\lambda_*}{b^m}\sum_{i=0}^{m}\left(\begin{array}{c} m \\ i \end{array}\right)\lambda_*^{m-i}(-a)^i \mu_{m-i}^{q*},
\end{aligned}$$

where we use a change of variable $y = (bz+a)/(1-\lambda)$ in the first step, and where $\lambda_* = 1-\lambda$ and $q_* = (bq+a)/(1-\lambda)$. We obtain

$$\begin{aligned}
E(\mathbf{1}(Z \le q)Z) &= \frac{\lambda_*}{b}(\lambda_*\mu_1^{q*} - a\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^2) &= \frac{\lambda_*}{b^2}(\lambda_*^2\mu_2^{q*} - 2a\lambda_*\mu_1^{q*} + a^2\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^3) &= \frac{\lambda_*}{b^3}(\lambda_*^3\mu_3^{q*} - 3a\lambda_*^2\mu_2^{q*} + 3a^2\lambda_*\mu_1^{q*} - a^3\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^4) &= \frac{\lambda_*}{b^4}(\lambda_*^4\mu_4^{q*} - 4a\lambda_*^3\mu_3^{q*} + 6a^2\lambda_*^2\mu_2^{q*} - 4a^3\lambda_*\mu_1^{q*} + a^4\mu_0^{q*}), \\
E(\mathbf{1}(Z \le q)Z^5) &= \frac{\lambda_*}{b^5}(\lambda_*^5\mu_5^{q*} - 5a\lambda_*^4\mu_4^{q*} + 10a^2\lambda_*^3\mu_3^{q*} - 10a^3\lambda_*^2\mu_2^{q*} + 5a^4\lambda_*\mu_1^{q*} - a^5\mu_0^{q*}).
\end{aligned}$$

Note that in the computation of ES we use $E\left(Z \mid Z \le q_a\right) = E(\mathbf{1}(Z \le q_a)Z)/\alpha$.

For $E\left(Z^5\right)$ we build on Jondeau and Rockinger (2003), who rely on the result of Gradshteyn and Ryzhik (1994):

$$\int_0^\infty x^{\mu-1}(p+qx^\nu)^{-(n+1)}dx = \frac{1}{\nu p^{n+1}}\left(\frac{p}{q}\right)^{\mu/\nu}\frac{\Gamma(\mu/\nu)\Gamma[1+n-(\mu/\nu)]}{\Gamma(1+n)}, \tag{A12}$$

where $0 < \mu/\nu < n+1$, $p \neq 0$, $q \neq 0$, $\Gamma(\cdot)$ is the gamma function with $\Gamma(x) = (x-1)\Gamma(x-1)$ and $\Gamma(1/2) = \sqrt{\pi}$.

Consider the variable $Y = Za + b$ with density

$$h\left(y\right) = \begin{cases} c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{if } y \leq 0, \\ c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{if } y > 0. \end{cases}$$

We have

$$\begin{aligned} E(Y^5) &= m_5 = \int_{-\infty}^0 y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} dy \\ &\quad + \int_0^\infty y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} dy \\ &= I_l + I_r, \end{aligned}$$

and on using (A12) we get

$$\begin{aligned} I_l &= \int_{-\infty}^0 y^5 c\left(1+\frac{1}{\nu-2}\left(\frac{y}{(1-\lambda)}\right)^2\right)^{-(\nu+1)/2} dy \\ &= c(1-\lambda)^6\int_{-\infty}^0 x^5\left(1+\frac{x^2}{\nu-2}\right)^{-(\nu+1)/2} dx \\ &= -c(1-\lambda)^6(\nu-2)^3\frac{\Gamma[(\nu-5)/2]}{\Gamma[(\nu+1)/2]} \\ &= -8c(1-\lambda)^6\frac{(\nu-2)^3}{(\nu-1)(\nu-3)(\nu-5)}, \end{aligned}$$

where we use the change of variable $x = y/(1-\lambda)$ in the first step. Similarly, $I_r = 8c(1+\lambda)^6(\nu-2)^3/[(\nu-1)(\nu-3)(\nu-5)]$ and $m_5 = 8c(\nu-2)^3[(1+\lambda)^6-(1-\lambda)^6]/[(\nu-1)(\nu-3)(\nu-5)]$. We then have

$$E\left(Z^5\right) = \frac{E(Y-a)^5}{b^5} = \frac{m_5+4a^5-5am_4-10a^3m_2+10a^2m_3}{b^5}.$$

Table A1: Proportions of cases when the ML estimator did not converge to a valid point or when the indicated approximation failed. The reported numbers are maxima taken over the considered confidence levels and horizons.

	$\beta = .4$				$\beta = .8$				$\beta = .89$			
	Root-k	G-C	S-W	Sim	Root-k	G-C	S-W	Sim	Root-k	G-C	S-W	Sim
					$T = 500$							
$\gamma = 0$	0.016	0.016	0.016	0.016	0.001	0.001	0.001	0.001	0.074	0.074	0.074	0.074
$\gamma = 0.1$	0.012	0.011	0.011	0.011	0.008	0.008	0.012	0.008	0.111	0.111	0.111	0.111
$\gamma = 0.2$	0.017	0.017	0.048	0.017	0.013	0.013	0.018	0.013	0.176	0.177	0.176	0.176
					$T = 1000$							
$\gamma = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.039	0.039	0.039	0.039
$\gamma = 0.1$	0.003	0.003	0.003	0.003	0.000	0.000	0.000	0.000	0.056	0.056	0.056	0.056
$\gamma = 0.2$	0.008	0.007	0.009	0.007	0.001	0.006	0.002	0.012	0.085	0.090	0.086	0.100

Table A2: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.1y_{t-1}^2 + 0.4h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.0011	0.1199	0.0744	-0.0178	0.3038	0.1487	-0.1385	0.2037	0.1169	-0.1560	0.4974	0.2339
G-C	-0.0016	0.0420	0.0438	-0.0050	0.0747	0.0784	0.0091	0.0924	0.0948	0.0085	0.1507	0.1616
S-W	-0.0027	0.0416	0.0443	-0.0052	0.0744	0.0794	-0.0062	0.0830	0.0879	-0.0015	0.1418	0.1541
Sim	-0.0069	0.0437	0.1017	-0.0108	0.0774	0.1531	-0.0129	0.0880	0.0996	-0.0104	0.1481	0.1756
						$T=1000$						
Root-k	0.0035	0.0727	0.0362	-0.0114	0.2031	0.0723	-0.1330	0.1282	0.0569	-0.1484	0.3375	0.1137
G-C	0.0055	0.0214	0.0213	0.0050	0.0363	0.0372	0.0146	0.0433	0.0445	0.0151	0.0695	0.0740
S-W	0.0027	0.0213	0.0215	0.0027	0.0363	0.0375	-0.0029	0.0407	0.0419	0.0022	0.0673	0.0715
Sim	0.0017	0.0219	0.0347	0.0018	0.0379	0.0672	-0.0055	0.0416	0.0437	-0.0012	0.0698	0.0741
	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2067	0.2756	0.1487	-0.2245	0.6516	0.2975	-0.4261	0.4825	0.1952	-0.4414	0.9796	0.3904
G-C	0.0288	0.1430	0.1469	0.0274	0.2194	0.2392	0.1364	0.3947	0.3974	0.1192	0.5137	0.5771
S-W	-0.0072	0.1154	0.1223	0.0025	0.1943	0.2127	-0.0129	0.2127	0.2277	0.0068	0.3405	0.3784
Sim	-0.0154	0.1221	0.1798	-0.0103	0.2034	0.3564	-0.0215	0.2295	0.2622	-0.0038	0.3615	0.4422
						$T=1000$						
Root-k	-0.2006	0.1793	0.0723	-0.2161	0.4476	0.1446	-0.4192	0.3543	0.0949	-0.4319	0.7102	0.1899
G-C	0.0316	0.0647	0.0668	0.0305	0.0986	0.1067	0.1236	0.1722	0.1737	0.1010	0.2168	0.2374
S-W	-0.0044	0.0565	0.0582	0.0046	0.0916	0.0983	-0.0153	0.1025	0.1065	-0.0001	0.1592	0.1715
Sim	-0.0081	0.0579	0.0804	-0.0024	0.0943	0.1702	-0.0191	0.1063	0.1123	-0.0057	0.1671	0.1793

Table A3: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.05y_{t-1}^2 + 0.4h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.0838	0.1872	0.1097	-0.1350	0.4496	0.2194	-0.3367	0.3904	0.1725	-0.4266	0.8522	0.3451
G-C	0.0176	0.0716	0.0729	0.0181	0.1215	0.1247	0.1072	0.2598	0.3028	0.1487	0.4707	0.5528
S-W	0.0156	0.0697	0.0698	0.0188	0.1179	0.1184	-0.0124	0.1599	0.1658	0.0026	0.2759	0.2982
Sim	-0.0155	0.0665	0.1029	-0.0162	0.1120	0.2201	-0.0341	0.1572	0.1607	-0.0218	0.2689	0.2823
						$T=1000$						
Root-k	-0.0686	0.0971	0.0535	-0.1134	0.2789	0.1069	-0.3175	0.2369	0.0841	-0.3994	0.5688	0.1681
G-C	0.0356	0.0324	0.0342	0.0397	0.0541	0.0564	0.1192	0.1226	0.1301	0.1453	0.2085	0.2197
S-W	0.0365	0.0324	0.0318	0.0459	0.0554	0.0527	0.0133	0.0718	0.0756	0.0287	0.1268	0.1342
Sim	0.0036	0.0282	0.0442	0.0070	0.0483	0.0834	-0.0007	0.0680	0.0766	0.0115	0.1208	0.1305
	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.4669	0.5678	0.2195	-0.5792	1.1866	0.4389	-0.8450	1.1638	0.2880	-0.9896	1.9949	0.5761
G-C	0.0504	0.2971	0.3325	0.0746	0.5136	0.5866	0.1843	0.6975	0.7549	0.2678	1.2226	1.3603
S-W	-0.0217	0.2353	0.2453	0.0007	0.4081	0.4494	-0.0880	0.4487	0.4903	-0.0515	0.7757	0.9147
Sim	-0.0442	0.2328	0.2990	-0.0274	0.3978	0.6056	-0.0664	0.4847	0.5137	-0.0239	0.8356	0.9082
						$T=1000$						
Root-k	-0.4451	0.3673	0.1069	-0.5486	0.8183	0.2139	-0.8199	0.8825	0.1403	-0.9644	1.5962	0.2807
G-C	0.0787	0.1343	0.1498	0.0955	0.2270	0.2507	0.2250	0.3690	0.3755	0.2835	0.6486	0.6632
S-W	0.0079	0.1058	0.1122	0.0282	0.1872	0.2034	-0.0583	0.2034	0.2218	-0.0357	0.3529	0.4087
Sim	-0.0025	0.1011	0.1398	0.0133	0.1798	0.2649	-0.0087	0.2170	0.2467	0.0235	0.3855	0.4247

Table A4: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.4h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.1550	0.1923	0.1061	-0.2367	0.5305	0.2122	-0.5240	0.5183	0.1669	-0.6903	1.1974	0.3338
G-C	0.0477	0.0711	0.0806	0.0560	0.1250	0.1394	0.2882	0.4914	0.5593	0.3983	0.9888	1.1183
S-W	0.0455	0.0659	0.0709	0.0596	0.1166	0.1222	0.0134	0.1689	0.1904	0.0463	0.3209	0.3637
Sim	-0.0137	0.0603	0.1004	-0.0099	0.1056	0.2191	-0.0321	0.1630	0.1866	-0.0133	0.2980	0.3359
						$T=1000$						
Root-k	-0.1521	0.1403	0.0570	-0.2310	0.4246	0.1140	-0.5204	0.4339	0.0896	-0.6835	1.0253	0.1792
G-C	0.0579	0.0385	0.0389	0.0681	0.0637	0.0655	0.2778	0.2611	0.2286	0.3516	0.4522	0.4170
S-W	0.0565	0.0367	0.0346	0.0734	0.0614	0.0581	0.0222	0.0863	0.0902	0.0470	0.1548	0.1704
Sim	-0.0030	0.0307	0.0491	0.0028	0.0514	0.1001	-0.0098	0.0804	0.0906	0.0054	0.1422	0.1596

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.7145	0.8124	0.2123	-0.9268	1.7687	0.4246	-1.2594	1.9530	0.2786	-1.5780	3.6402	0.5573
G-C	0.1230	0.3903	0.4758	0.1865	0.7622	0.9170	0.2453	0.7538	0.8012	0.4029	1.5126	1.7512
S-W	0.0069	0.2560	0.2884	0.0533	0.4946	0.5656	-0.0916	0.5463	0.6246	-0.0280	1.0632	1.2839
Sim	-0.0421	0.2466	0.3424	-0.0202	0.4505	0.7434	-0.0635	0.5694	0.6528	-0.0129	1.0621	1.2116
						$T=1000$						
Root-k	-0.7105	0.7044	0.1140	-0.9194	1.5472	0.2279	-1.2550	1.8076	0.1496	-1.5695	3.3420	0.2686
G-C	0.1327	0.1923	0.2091	0.1751	0.3377	0.3779	0.3256	0.4776	0.4153	0.4853	0.9736	0.8606
S-W	0.0160	0.1301	0.1361	0.0491	0.2352	0.2652	-0.0913	0.2765	0.2866	-0.0605	0.4948	0.5834
Sim	-0.0145	0.1209	0.1696	0.0048	0.2130	0.3516	-0.0212	0.2766	0.3143	0.0151	0.4975	0.5728

Table A5: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.1y_{t-1}^2 + 0.8h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	0.0109	0.1028	0.0831	0.0070	0.2943	0.1661	-0.1884	0.1876	0.1306	-0.2928	0.5099	0.2613
G-C	-0.0044	0.0813	0.0818	-0.0030	0.1737	0.1740	0.0180	0.1662	0.1672	0.0242	0.3682	0.3809
S-W	-0.0075	0.0832	0.0813	-0.0079	0.1759	0.1729	-0.0184	0.1574	0.1538	-0.0314	0.3417	0.3417
Sim	-0.0067	0.0811	0.0950	-0.0090	0.1739	0.2285	-0.0157	0.1543	0.1617	-0.0328	0.3393	0.3540
						$T=1000$						
Root-k	0.0182	0.0495	0.0388	0.0172	0.1780	0.0775	-0.1792	0.1014	0.0610	-0.2800	0.3230	0.1219
G-C	0.0055	0.0367	0.0390	0.0104	0.0782	0.0838	0.0309	0.0745	0.0792	0.0473	0.1666	0.1807
S-W	0.0009	0.0365	0.0388	0.0036	0.0776	0.0831	-0.0053	0.0690	0.0734	-0.0088	0.1515	0.1640
Sim	0.0028	0.0368	0.0464	0.0044	0.0795	0.1155	-0.0008	0.0695	0.0751	-0.0073	0.1550	0.1669

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2895	0.2741	0.1662	-0.4416	0.7194	0.3323	-0.6028	0.6029	0.2181	-0.9204	1.4810	0.4362
G-C	0.0540	0.2519	0.2527	0.0798	0.5654	0.5988	0.2730	0.6877	0.6274	0.3881	1.5960	1.5810
S-W	-0.0269	0.2143	0.2090	-0.0442	0.4660	0.4694	-0.0319	0.3645	0.3588	-0.0725	0.8212	0.8378
Sim	-0.0202	0.2117	0.2439	-0.0444	0.4656	0.5897	-0.0297	0.3627	0.3798	-0.0770	0.8245	0.8724
						$T=1000$						
Root-k	-0.2791	0.1635	0.0775	-0.4271	0.4798	0.1551	-0.5909	0.4545	0.1018	-0.9036	1.1606	0.2035
G-C	0.0663	0.1127	0.1172	0.1049	0.2564	0.2760	0.2824	0.3416	0.2901	0.4249	0.8170	0.7314
S-W	-0.0108	0.0942	0.0998	-0.0153	0.2065	0.2253	-0.0111	0.1604	0.1706	-0.0292	0.3672	0.3996
Sim	-0.0018	0.0955	0.1162	-0.0118	0.2127	0.2907	-0.0064	0.1631	0.1767	-0.0298	0.3805	0.4101

Table A6: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.8h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2107	0.1833	0.1250	-0.4060	0.6224	0.2500	-0.7258	0.7103	0.1966	-1.3422	2.3136	0.3932
G-C	0.0863	0.1554	0.1625	0.1628	0.3866	0.3752	0.6075	1.2142	1.0112	1.5996	6.4715	3.7075
S-W	0.0678	0.1431	0.1495	0.0945	0.3237	0.3209	0.0306	0.3088	0.3456	0.0573	0.8202	0.8742
Sim	-0.0010	0.1328	0.1537	-0.0045	0.3022	0.3967	-0.0041	0.3045	0.3436	-0.0158	0.7958	0.8549
						$T=1000$						
Root-k	-0.2093	0.1110	0.0585	-0.4046	0.4602	0.1170	-0.7237	0.6019	0.0920	-1.3397	2.0751	0.1840
G-C	0.0857	0.0716	0.0753	0.1686	0.1870	0.1741	0.5643	0.6909	0.4118	1.5103	4.1453	1.8562
S-W	0.0668	0.0645	0.0694	0.1002	0.1464	0.1485	0.0292	0.1397	0.1606	0.0643	0.3725	0.4050
Sim	-0.0002	0.0572	0.0742	0.0008	0.1324	0.1707	0.0003	0.1351	0.1601	-0.0009	0.3621	0.3976
	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.9982	1.2252	0.2501	-1.8291	3.9333	0.5001	-1.7362	3.3279	0.3282	-3.1880	10.7864	0.6564
G-C	0.2453	0.7030	0.7574	0.6614	2.9366	2.5159	0.2630	0.9006	0.8973	-0.5548	3.8417	2.6051
S-W	0.0170	0.4457	0.4985	0.0479	1.2109	1.2534	-0.0789	0.8390	0.9629	-0.0686	2.6008	2.6264
Sim	-0.0055	0.4471	0.6427	-0.0221	1.1853	2.2368	-0.0070	0.8701	1.0008	-0.0311	2.5906	2.5881
						$T=1000$						
Root-k	-0.9955	1.0886	0.1170	-1.8261	3.6370	0.2341	-1.7330	3.1595	0.1536	-3.1839	10.4207	0.3072
G-C	0.2342	0.3353	0.3322	0.6133	1.4107	1.1900	0.3805	0.4990	0.3990	-0.0732	0.8876	1.1909
S-W	0.0162	0.2033	0.2319	0.0578	0.5530	0.6053	-0.0822	0.3934	0.4457	-0.0642	1.1934	1.3070
Sim	0.0010	0.1980	0.2849	-0.0017	0.5413	1.1025	0.0045	0.3937	0.4653	-0.0016	1.1907	1.2980

Table A7: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.1y_{t-1}^2 + 0.89h_{t-1}$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.0115	0.0560	0.0689	-0.0393	0.1260	0.1378	-0.2017	0.1257	0.1083	-0.3758	0.3127	0.2167
G-C	-0.0211	0.0669	0.0822	-0.0346	0.1750	0.2081	0.0039	0.1290	0.1633	0.0157	0.3603	0.4520
S-W	-0.0257	0.0668	0.0818	-0.0428	0.1747	0.2066	-0.0351	0.1221	0.1522	-0.0667	0.3286	0.4029
Sim	-0.0234	0.0675	0.0873	-0.0395	0.1768	0.2492	-0.0281	0.1234	0.1551	-0.0555	0.3339	0.4119
						$T=1000$						
Root-k	0.0198	0.0330	0.0346	0.0102	0.0842	0.0692	-0.1703	0.0705	0.0544	-0.3268	0.1933	0.1089
G-C	0.0120	0.0335	0.0424	0.0203	0.0860	0.1103	0.0441	0.0693	0.0835	0.0881	0.1977	0.2352
S-W	0.0073	0.0331	0.0422	0.0114	0.0846	0.1095	0.0058	0.0611	0.0783	0.0055	0.1621	0.2121
Sim	0.0100	0.0336	0.0445	0.0155	0.0865	0.1272	0.0121	0.0612	0.0799	0.0173	0.1693	0.2177

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2996	0.2002	0.1378	-0.5474	0.5160	0.2756	-0.5926	0.5135	0.1809	-1.0700	1.4624	0.3618
G-C	0.0402	0.1922	0.2389	0.0987	0.5776	0.7176	0.2730	0.5406	0.5452	0.5524	1.8834	1.6727
S-W	-0.0443	0.1655	0.2048	-0.0863	0.4442	0.5483	-0.0412	0.2658	0.3358	-0.0943	0.7568	0.9510
Sim	-0.0305	0.1679	0.2207	-0.0652	0.4562	0.6196	-0.0327	0.2701	0.3459	-0.0790	0.7804	0.9867
						$T=1000$						
Root-k	-0.2670	0.1245	0.0692	-0.4964	0.3522	0.1385	-0.5656	0.4119	0.0909	-1.0291	1.2260	0.1818
G-C	0.0842	0.1087	0.1205	0.1793	0.3421	0.3620	0.3224	0.3790	0.2685	0.6703	1.4774	0.8792
S-W	0.0020	0.0827	0.1052	-0.0025	0.2216	0.2883	0.0110	0.1378	0.1713	0.0070	0.3908	0.4680
Sim	0.0143	0.0837	0.1158	0.0204	0.2362	0.3253	0.0168	0.1379	0.1770	0.0203	0.4172	0.4872

Table A8: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.1y_{t-1}^2 + 0.89h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.1132	0.0765	0.0902	-0.2594	0.1970	0.1803	-0.4373	0.3048	0.1418	-0.8930	1.0280	0.2835
G-C	0.0187	0.0835	0.1229	0.0478	0.2206	0.3259	0.1861	0.3020	0.3889	0.5968	1.5351	1.6512
S-W	0.0103	0.0815	0.1176	0.0130	0.2059	0.2989	-0.0237	0.1645	0.2394	-0.0461	0.4612	0.6861
Sim	-0.0218	0.0778	0.1148	-0.0434	0.1978	0.3194	-0.0301	0.1619	0.2431	-0.0657	0.4546	0.6858
						$T=1000$						
Root-k	-0.1040	0.0490	0.0519	-0.2450	0.1373	0.1038	-0.4406	0.2868	0.0816	-0.9030	1.0280	0.1632
G-C	0.0363	0.0491	0.0697	0.0886	0.1341	0.1864	0.2058	0.1900	0.1951	0.6565	1.2642	0.8725
S-W	0.0280	0.0479	0.0675	0.0528	0.1213	0.1740	-0.0014	0.0888	0.1327	0.0083	0.2519	0.3827
Sim	-0.0056	0.0459	0.0697	-0.0087	0.1120	0.1847	-0.0076	0.0891	0.1350	-0.0113	0.2498	0.3840

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.6096	0.5333	0.1803	-1.2260	1.8457	0.3607	-1.0768	1.4609	0.2367	-2.1433	5.3391	0.4734
G-C	0.0921	0.2994	0.4251	0.2775	1.1214	1.5135	0.2730	0.6161	0.7021	0.2138	1.2817	1.6140
S-W	-0.0421	0.2317	0.3325	-0.0770	0.6711	0.9824	-0.1008	0.3977	0.5697	-0.1792	1.2675	1.8301
Sim	-0.0360	0.2321	0.3544	-0.0770	0.6750	1.0788	-0.0416	0.4118	0.6136	-0.1021	1.3057	1.8294
						$T=1000$						
Root-k	-0.6182	0.5312	0.1038	-1.2471	1.9229	0.2075	-1.1066	1.5959	0.1362	-2.1710	5.4702	0.2724
G-C	0.1220	0.1709	0.2238	0.3507	0.7290	0.7718	0.3375	0.4355	0.3927	0.4337	0.9015	0.8487
S-W	-0.0157	0.1233	0.1826	-0.0123	0.3606	0.5431	-0.0722	0.2079	0.3043	-0.1011	0.6726	0.9793
Sim	-0.0094	0.1237	0.1953	-0.0102	0.3668	0.5828	-0.0083	0.2178	0.3290	-0.0135	0.7084	1.0300

Table A9: Simulations results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where $\varepsilon_t \sim nid(0,1)$ and $h_t = \omega + 0.89h_{t-1} + 0.2\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. MSE is the mean square error and EAV is the average estimated asymptotic variance.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.2120	0.1029	0.0798	-0.4592	0.3233	0.1596	-0.6426	0.5654	0.1255	-1.3467	2.1975	0.2509
G-C	0.0362	0.0921	0.1233	0.0945	0.2534	0.3400	0.4337	0.6265	0.5515	1.5272	5.2704	2.9383
S-W	0.0180	0.0849	0.1145	0.0126	0.2053	0.2928	-0.0266	0.1839	0.2542	-0.0508	0.5295	0.6991
Sim	-0.0326	0.0797	0.1163	-0.0608	0.1999	0.3250	-0.0458	0.1778	0.2536	-0.0907	0.5219	0.6952
						$T=1000$						
Root-k	-0.2061	0.0955	0.0566	-0.4515	0.3160	0.1132	-0.6570	0.6901	0.0890	-1.3453	2.1729	0.1780
G-C	0.0551	0.0515	0.0813	0.1396	0.1528	0.2229	0.4822	0.6784	0.3435	1.6525	5.0266	1.6592
S-W	0.0358	0.0490	0.0772	0.0510	0.1197	0.1964	-0.0031	0.0930	0.1626	0.0034	0.2669	0.3894
Sim	-0.0179	0.0481	0.0739	-0.0289	0.1152	0.2004	-0.0241	0.0940	0.1626	-0.0393	0.2653	0.3910

	$VaR_{T,k}^{0.99}$						$ES_{T,k}^{0.99}$					
	$k=5$			$k=10$			$k=5$			$k=10$		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
						$T=500$						
Root-k	-0.8730	1.0125	0.1596	-1.8145	3.9774	0.3192	-1.4776	2.8100	0.2095	-3.0310	10.8171	0.4189
G-C	0.1408	0.4116	0.4563	0.5559	2.3048	1.7812	0.1683	0.5053	0.5514	-1.0283	3.2347	1.9745
S-W	-0.0479	0.2668	0.3608	-0.0844	0.7946	0.9891	-0.1337	0.4830	0.6028	-0.2150	1.3649	2.0271
Sim	-0.0533	0.2646	0.4266	-0.1061	0.8003	2.1882	-0.0662	0.4831	0.6252	-0.1596	1.3606	2.0313
						$T=1000$						
Root-k	-0.8986	1.2706	0.1132	-1.8223	4.0235	0.2264	-1.4904	2.9020	0.1486	-3.0539	10.9486	0.2971
G-C	0.1755	0.2347	0.3026	0.5999	1.3016	0.9695	0.2655	0.3389	0.2949	-0.7958	1.7170	0.9450
S-W	-0.0217	0.1316	0.2278	-0.0199	0.3914	0.5297	-0.1021	0.2414	0.4031	-0.1153	0.8276	1.0569
Sim	-0.0293	0.1327	0.2678	-0.0427	0.3975	1.4240	-0.0309	0.2392	0.4202	-0.0539	0.8587	1.0810

Table A10: Simulation results for data generated according to $y_t = \sqrt{h_t}\varepsilon_t$, where ε_t is skew-t distributed with $\lambda = -0.2$ and $\nu = 8$ and $h_t = \omega + 0.05y_{t-1}^2 + 0.8h_{t-1} + 0.1\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$. The sample size and the horizon is set to 1000 observations and 5 periods, respectively. MSE is the mean square error and EAV is the average estimated asymptotic variance. The averages are for the true values.

	$VaR_{T,k}^{0.95}$						$ES_{T,k}^{0.95}$					
Average	4.8503						6.8472					
	QML			ML			QML			ML		
Method	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV	Bias	MSE	EAV
Root-k	-0.1992	0.1715	0.1703	0.0356	0.1675	0.0473	-1.0146	1.1805	0.2428	-0.0787	0.3552	0.0982
G-C	-0.0602	0.1127	0.1400	0.1964	0.1640	0.0667	-0.3302	0.4342	0.4365	1.9145	6.4573	2.8491
S-W	-0.0676	0.1113	0.1275	0.0969	0.1084	0.0965	-0.5536	0.5257	0.2214	0.1070	0.3099	0.1507
Sim	-0.1077	0.1142	0.1479	0.0175	0.0922	0.1158	-0.5647	0.5364	0.2698	0.0171	0.2667	0.2046

References

Andreev, A., Kanto, A., 2005. Conditional value-at-risk estimation using non-integer values of degrees of freedom in Student's t-distribution. Journal of Risk 7, 55–61.

Baillie, R. T., Bollerslev, T., 1992. Prediction in dynamic models with time-dependent conditional variances. Journal of Econometrics 52, 91–113.

Barton, D. E., Dennis, K. E., 1952. The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 39, 425–427.

Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? Journal of Finance 57, 1093 – 1111.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Wooldridge, J. M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Boudoukh, J., Richardson, M., Whitelaw, R. F., 1997. Investigation of a class of volatility estimators. Journal of Derivatives 4, 63–71.

Breuer, T., Jandačka, M., 2007. Temporal aggregation of GARCH models: Conditional kurtosis and optimal frequency. Working paper, PPE Research Centre, University of Applied Sciences Vorarlberg.

Brummelhuis, R., Guégan, D., 2005. Multiperiod conditional distribution functions for conditionally normal GARCH(1,1) models. Journal of Applied Probability 42, 426–445.

Brummelhuis, R., Kaufmann, R., 2007. Time scaling of VAR in GARCH(1,1) and AR(1)-GARCH(1,1) processes. Journal of Risk 9, 39–94.

Brännäs, K., Nordman, N., 2003. Conditional skewness modelling for stock returns. Applied Economics Letters 10, 725–728.

Chan, N. H., Deng, S.-J., Peng, L., Xia, Z., 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics 137, 556–576.

Chen, S. X., 2008. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 6, 87–107.

Chen, S. X., Tang, C. Y., 2005. Nonparametric inference of value-at-risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

Christoffersen, P., Diebold, F. X., 2000. How relevant is volatility forecasting for financial risk management? The The Review of Economics and Statistics 82, 12–22.

Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. Journal of Risk 7, 1–28.

Christoffersen, P., Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics 2, 84–108.

Cornish, E. A., Fisher, R. A., 1938. Moments and cumulants in the specification of distributions. Review of the International Statistical Institute 5, 307–320.

Cotter, J., 2007. Varying the VaR for unconditional and conditional environments. Journal of International Money and Finance 26, 1338–1354.

Diebold, F. X., Hickman, A., Inoue, A., Schuermann, T., 1997. Converting 1-day volatility to h-day volatility: Scaling by Root-h is worse than you think. Working paper 97-34, Wharton Financial Institutions Center.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Dowd, K., 2007. Validating multiple-period density-forecasting models. Journal of Forecasting 26, 251–270.

Duan, J.-C., Gauthier, G., Simonato, J.-G., Sasseville, C., 2006. Approximating the GJR-GARCH and EGARCH option pricing models analytically. Journal of Computational Finance 9, 41–69.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F., 1990. Stock volatility and the crash of '87: Discussion. Review of Financial Studies 3, 103–106.

Engle, R. F., Ng, V. K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Giamouridis, D., 2006. Estimation risk in financial risk management: A correction. Journal of Risk 8, 121–125.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gradshteyn, I. S., Ryzhik, I. M., 1994. Tables of Integrals, Series and Products, 5th Edition. Academic Press, New York.

Hafner, C. M., 2003. Fourth moment structure of multivariate GARCH models. Journal of Financial Econometrics 1, 26–54.

Hafner, C. M., 2008. Temporal aggregation of multivariate GARCH processes. Journal of Econometrics 142, 467–483.

Hansen, B. E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. Journal of Econometrics 135, 377–398.

Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34, 465–487.

He, C., Teräsvirta, T., 1999. Properties of moments of a family of GARCH processes. Journal of Econometrics 92, 173–192.

Jondeau, E., Rockinger, M., 2001. Gram-Charlier densities. Journal of Economic Dynamics and Control 25, 1457–1483.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27, 1699–1737.

Jondeau, E., Rockinger, M., 2006. The Copula-GARCH model of conditional dependencies: An international stock market application. Journal of International Money and Finance 25, 827–853.

Jorion, P., 2007. Value at Risk: The New Benchmark for Managing Financial Risk, 3rd Edition. McGraw-Hill, New York.

Kaibila, P., He, Z., 2004. The adjustment of prediction intervals to account for errors in parameter estimation. Journal of Time Series Analysis 25, 351–358.

Lan, Y., Hu, Z. M., Johnson, J., 2007. Can value-at-risk be made more valuable? Resampling techniques for risk uncertainty. Working paper, Financial Studies Discipline, UWA Business School, The University of Western Australia.

Ling, S., McAleer, M., 2002. Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106, 109–117.

Manistre, B. J., Hancock, G. H., 2005. Variance of the CTE estimator. North American Actuarial Journal 9, 129–156.

Marcellino, M., Stock, J. H., Watson, M. W., 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135, 499–526.

McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance 7, 271–300.

Pesaran, M. H., Pick, A., Timmermann, A., 2009. Variable selection and inference for multi-period forecasting problems. Cambridge working papers in economics, Faculty of Economics, University of Cambridge.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P., 2007. Numerical Recipes: The Art of Scientific Computing, 3rd Edition. Cambridge University Press, New York.

Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115–129.

Schmidt, P., 1974. The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. Econometrica 42, 303–309.

Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies 62, 639–661.

Taylor, J. W., 1999. A quantile regression approach to estimating the distribution of multiperiod returns. Journal of Derivatives 7, 64–78.

Taylor, J. W., 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 19, 299–311.

Theodossiou, P., 1998. Financial data and the skewed generalized t distribution. Management Science 44, 1650–1661.

Wong, C.-M., So, M. K. P., 2003. On conditional moments of GARCH models, with applications to multiple period value at risk estimation. Statistica Sinica 13, 1015–1044.

Wong, C.-M., So, M. K. P., 2007. On the estimation of multiple period expected shortfall and median shortfall using GARCH models. Working paper, Hong Kong University of Science and Technology.