Time Scaling for GARCH(1,1) and AR(1)-GARCH(1,1) Processes

Raymond Brummelhuis
Birkbeck College
University of London
School of Economics, Mathematics, and Statistics
Malet Street, London WC1E 7HX
UK
r.brummelhuis@statistics.bbk.ac.uk

Roger Kaufmann
RiskLab
Department of Mathematics
ETH Zurich
CH–8092 Zürich
Switzerland
kaufmann@math.ethz.ch

November 7, 2004

Abstract. This paper investigates the estimation of a 10-day value-at-risk based on a data set of 250 daily values. The commonly used square-root-of-time rule, which scales the 1-day 99% value-at-risk with a factor $\sqrt{10}$, is compared with alternative 10-day estimators in the case of random walks, GARCH(1,1) and AR(1)-GARCH(1,1) processes. Additionally, some theoretical results on N-day value-at-risk in such models are presented. The overall conclusion is that, although not perfect, the $\sqrt{10}$-rule performs remarkably well.

1 Introduction

This paper addresses the following fundamental question of present-day financial risk management: *How should one scale a short term value-at-risk to get a long term value-at-risk?* More precisely we are interested in the following problem: *What is the best estimation for an unconditional 10-day 99% value-at-risk if only data for 250 trading days are available? Can the 1-day 99% value-at-risk simply be re-scaled by* $\sqrt{10}$ *as in the case*

Key words and phrases. Value-at-risk, scaling rules, random walk, GARCH(1,1) process, AR(1)-GARCH(1,1) process.

Research of the second author was supported by Credit Suisse Group, Swiss Re and UBS AG through RiskLab, Switzerland.

of independent normally distributed returns? An important reason why banks and other regulated financial companies are interested in this square-root-of-time rule – and, more generally, in estimating a 10-day risk – is the Basle Committee on Banking Supervision's recommendation [2] that such companies should have a capital reserve on their investment portfolio proportional to the 10-day value-at-risk at the 99% confidence level. Recently (summer 2004), this recommendation has become law within the European Union. The Basle Committee allows the use of the square-root-of-time rule to get a 10-day 99% value-at-risk number out of a one-day 99% value-at-risk estimate, but will back-test the bank's prediction against actually occurred losses, and impose a penalty if the back-tested value-at-risk exceeds the predicted one. On the other hand, banks will want to limit their capital reserves, and will therefore be concerned to not over-estimate their value-at-risk. In this paper, we investigate whether the simple square-root-of-time rule is justified, and whether easily implemented better performing alternatives are available. Earlier work on this topic (but focussing on volatility fluctuations rather than value-at-risk) includes Christoffersen et al. [7] and Diebold et al. [8].

The reason for the restriction to a 250-day sample of return data may not be so clear at first, but this is an apparent need felt on the practitioners' side; indeed, the problem was originally posed to us in these terms. As an explanation, one may think of situations where the entire portfolio is modelled by some single factor model and where, typically, longer time-series for the portfolio in question are not available, or where the portfolio's composition will have substantially changed over time so as to affect the stationarity assumption over longer periods.

The question we are addressing clearly does not have an absolute and universally valid answer, but is highly model-dependent. We will analyze the $\sqrt{10}$-rule for some of the models popularly used in financial risk-management practice: simple random walks and, more generally, GARCH(1,1) and AR(1)-GARCH(1,1) models, with typical market parameters[1]. For each of these models we will examine the validity of the $\sqrt{10}$-rule on small data sets, and compare with potential alternatives, both from an empirical and a theoretical point of view. As regards the former, we propose a number of alternative 10-day value-at-risk estimators and evaluate their performance, relative to the square root of time rule, on large numbers of 250-day samples, for each of the three types of time-series models mentioned above. As we will see, given the relative smallness of the sample, variability of the estimators will be as much of an issue as consistency.

As for the theoretical approach, if one assumes that 250 days is still a reasonably-sized sample to fit any of the above models for the one-day returns (an assumption which would certainly merit closer examination, but which one could be inclined to make for want of an alternative), then the question becomes that of how to convert, for each of these models, a one-day quantile into a quantile for the aggregated 10-day returns[2]. Mathematically,

[1] It is of course debatable whether restricting ourselves to these processes is reasonable, and whether other classes of models, in particular those including leverage effects, would not be more realistic. On the other hand, the models used here can with some justification be considered to be the 'working horses' of quantitative finance, and as such are a reasonable choice for an exploratory study as ours. See however Stărică [15] for severe criticism of the GARCH(1,1), both for long term and intermediate-term modelling. Also, conspicuously missing from our list are the Stochastic Volatility Models, but for these see Kaufmann [11], Chapter 3.

[2] Assuming of course the one-day return model is correctly specified and still relevant on a 10-day horizon; this point was forcefully made in Christoffersen et al. [7] and Diebold et al. [8]. The

this becomes a surprisingly delicate question once we leave the setting of i.i.d. normally distributed (or, more generally, stably distributed) returns. The basic reason is that one finds oneself in between two different asymptotic regimes which a priori are relevant: that of a confidence level α close to one and that of a relatively large sum of random variables. Take for example the case of a simple random walk $Z_N = \sum_{j=1}^{N} X_j$ with i.i.d. but heavy-tailed innovations X_j, whose tail distributions decay as Cx^{-k} for some positive k. By a well-known theorem of Feller, the same will be true for the tail distributions of Z_N (with a different, N-dependent, constant C). On the other hand, by the Central Limit Theorem or CLT, the center of the distribution of Z_N will, for sufficiently big N, be approximately normal, and the question becomes which of the two effects dominates for $N = 10$ and $\alpha = 0.99$. In Section 5 we will perform a detailed numerical study in the case of Student distributed X_j which will show that, for these values of N and α, the CLT-effect is the more important, though corrections are necessary.

A similar analysis can be carried out for the GARCH(1,1), although the mathematics involved is considerably more sophisticated. For example, Brummelhuis and Guégan [6] analyse the conditional tail probabilities of Z_N when $(X_j)_j$ is a normal GARCH(1,1), and show that the $\sqrt{N}$-rule for the α quantile is strongly violated as α tends to 1; cf. Brummelhuis [5] for extensions to GARCH models with Pareto-tailed innovations. It turns out, however, that for $N = 10$, Basel's 99% confidence level is not sufficiently close to 1 to make this result relevant. As we will see in this paper, a better qualitative understanding of the 10-day value-at-risk can again be obtained from the Central Limit Theorem, for which we will now need an appropriate martingale version. The quantitative agreement of the CLT-prediction turns out to be less than perfect, but we will propose a correction which, in the cases we considered, outperforms the $\sqrt{10}$-rule for both GARCH(1,1) and AR(1)-GARCH(1,1) processes. We note that in Christoffersen et al. [7] and Diebold et al. [8], the Drost-Nijman formulas [9] for weak GARCH(1,1) processes were used to scale-up the entire process from the 1-day to the 10-day scale: this is clearly relevant if one is foremost interested in the conditional volatitity. Contrary however to [7], [8], we do not consider the difference between weak GARCH and GARCH to be merely technical: on a very basic level, there is not even an unequivocal relationship between volatility and 1-day value-at-risk in a weak-GARCH(1,1) model, and it is far from clear which of the results from this paper or from Brummelhuis and Guégan [6] remain valid for weak-GARCH processes.

As mentioned, we will mostly concentrate on *unconditional* 10-day value-at-risk estimates, by which we will mean, more precisely, the value-at-risk associated to the strongly stationary return processes, for the various models we consider. In practice, unconditional risk measures are at least as important as conditional ones, if only because of the fact that current back-testing procedures used by regulators test unconditional value-at-risk. McNeil and Frey [13] concentrate on estimating *conditional* risk measures. For conditional quantiles, applying the square-root-of-time scaling is clearly outperformed by the Monte Carlo method presented in their paper, and tends to underestimate conditional quantiles. For confidence-parameter tending to 1, this is theoretically confirmed by the results of Brummelhuis and Guégan [6].

---

latter reference contains a general discussion of horizon problems in volatility forecastability, and finds a maximal horizon of 10–15 days, depending on the asset class.

We also stress that, for the various models, we assume that there is no trend in the data. Non-zero trends, even if negligible on a 1-day scale, can become important on a 10-day scale, since their contribution to the value-at-risk number grows linearly in time; cf. Section 6 below for a numerical illustration. Data should therefore always be de-trended. Similarly, the effect of a non-zero AR(1)-component may be negligible on a 1-day scale, but aggregation will increase its effect on longer time scales (although not linearly, as for trends), as will also be illustrated in Section 6. These two effects will in general immediately invalidate the $\sqrt{10}$-rule for normal random walks. Effects due to heteroscedasticity or heavy tails are generally more subtle, and can interfere with the former in complicated ways.

We end this introduction by giving a graphical illustration of the problems faced. As a reasonable model for daily log-returns $(X_t)_{t\in\mathbb{Z}}$, we take the following GARCH(1,1) process:

$$\begin{aligned} X_t &= \sigma_t\,\epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \\ \sigma_t^2 &= a_0 + a\,X_{t-1}^2 + b\,\sigma_{t-1}^2, \\ a_0 &= 3\cdot 10^{-6},\ a = 0.05,\ b = 0.92. \end{aligned}$$

We simulate 250 values $(X_t)_{t=1,\ldots,250}$ from this process. The top graph in Figure 1 shows the 10-day log-returns $(\sum_{t=10k-9}^{10k} X_t)_{k=1,\ldots,25}$ for non-overlapping periods. Obviously such a data set of 25 values is much too small to provide a reliable estimate of a 99% quantile. One possibility to increase the number of data is to use the overlapping values $(\sum_{t=k-9}^{k} X_t)_{k=10,11,\ldots,250}$. This is done in the middle graph in Figure 1. The main disadvantage with these overlapping data is that subsequent 10-day log-returns are strongly dependent, which is not desirable for doing statistics. An alternative method to increase the number of data is to first estimate a risk measure on daily basis, and then scale it with the square root of time (here $\sqrt{10}$) to get a two-week risk measure. This proceeding corresponds to scaling the daily log-returns by $\sqrt{10}$ to get values which can be interpreted as being representative for a two-week horizon, and then estimating the risk measure on these scaled values. The values $(\sqrt{10}\,X_t)_{t=1,\ldots,250}$ obtained via scaling are shown in the third graph in Figure 1. The disadvantage of this method is the fact that artificial instead of real 10-day log-returns are used.

The remainder of the paper is organized as follows: in Section 2 we review the three types of models we will be using in this study, and look at how the $\sqrt{10}$-rule performs on large data sets. Section 3 introduces a number of potential empirical and semi-empirical estimators for 99% quantiles over a 10-day horizon, whose performance relative to the $\sqrt{10}$-rule will be investigated in Section 4. In Section 5 we take a look at the interplay between tail asymptotics and the Central Limit Theorem, and discuss the relevance of the choice of confidence parameter. Section 6 presents some theoretical results for large-horizon quantile estimation in GARCH(1,1) and AR(1)-GARCH(1,1) models, and compares these once more with the $\sqrt{10}$-rule. The more technical proofs are collected in Appendix A and figures in Appendix B.

2 Price process models

For a given (daily) price process $(P_t)_{t\in\mathbb{Z}}$, one can alternatively consider the log-returns $X_t = \log(P_t/P_{t-1})$ or the percentage returns $Y_t = (P_t - P_{t-1})/P_{t-1}$. Log-returns have the nice property that N-day log-returns, which we will denote by Z_t^N or more briefly by Z_t, are simply the arithmetic sum of 1-day log-returns:

$$Z_t^N = \log\left(P_t/P_{t-N}\right) = X_t + X_{t-1} + \cdots + X_{t-(N-1)},$$

while percentage returns allow for a simple conversion between the percentage value-at-risk and the value-at-risk in monetary terms, by multiplication by present price. Assuming strong stationarity of the return process, the value-at-risk of the daily log-returns at the confidence level α is defined as the negative $(1-\alpha)$ quantile of $X_t \stackrel{d}{=} X$:

$$\mathrm{VaR}_\alpha(X) = -\inf\{x \in \mathbb{R} \mid \mathbb{P}[X_t \le x] \ge 1-\alpha\},$$

and similarly for $(Y_t)_t$ and $(Z_t^N)_t$. Since we are interested in the relation between 1-day and N-day value-at-risk, we will use log-returns. The two processes, and their value-at-risks, are in any case simply related by:

$$Y_t = e^{X_t} - 1, \quad \mathrm{VaR}_\alpha(Y) = 1 - \exp(-\mathrm{VaR}_\alpha(X)).$$

We will often simply write VaR_α^N for the N-day value-at-risk $\mathrm{VaR}_\alpha(Z_t^N)$, and VaR_α^1 for $\mathrm{VaR}_\alpha(X_t)$, the 1-day one.

As mentioned in the Introduction, we will work with three often-used classes of processes for modelling log-returns. The first one consists of the random walks models, where the daily log-returns $(X_t)_{t\in\mathbb{Z}}$ are assumed to be i.i.d.:

$$X_t = \sigma\,\epsilon_t, \quad \mathbb{E}[\epsilon_t] = 0,\ \mathbb{E}[\epsilon_t^2] = 1,\ \epsilon_t \text{ i.i.d.}, \tag{2.1}$$

with either normal, Student-t_8 or Student-t_4 innovations ϵ_t. Note that we will always normalise our Student-t distributions to have unit variance (instead of $\nu/(\nu-2)$).

The second, more flexible class of models is that of the GARCH(1,1) processes. The daily log-returns $(X_t)_{t\in\mathbb{Z}}$ in such models satisfy the recursive equation

$$\begin{aligned} &X_t = \sigma_t\,\epsilon_t, \quad \mathbb{E}[\epsilon_t] = 0,\ \mathbb{E}[\epsilon_t^2] = 1,\ \epsilon_t \text{ i.i.d.},\\ &\sigma_t^2 = a_0 + a\,X_{t-1}^2 + b\,\sigma_{t-1}^2. \end{aligned} \tag{2.2}$$

We will always assume second-order stationarity conditions: $a_0, a, b \ge 0$ and $a + b < 1$, which will also imply strong stationarity[3]. For real financial data, a is usually below 0.25 and b over 0.70, see e.g. Alexander [1]. Unless mentioned otherwise, we will fix the GARCH parameters as $a_0 = 3\cdot 10^{-6}$, $a = 0.05$ and $b = 0.92$, these being typical values for GARCH models fitted to foreign exchange rates, stock indices, single stocks and 10-year government bonds. As for the random walks (which are anyhow included amongst the GARCH(1,1) as a simple special case), we assume either normal, Student-t_8 or Student-t_4 innovations ϵ_t.

[3] A sufficient condition for the existence of a unique strongly stationary GARCH(1,1) is that $\mathbb{E}\left(\log(a_1\epsilon^2 + b_1)\right) < 0$; cf. e.g. Basrak et al. [3] and its references.

Our third, and final, family of models is the class of AR(1)-GARCH(1,1) processes, in which the mean is modelled by a first order auto-regressive AR(1), with a GARCH(1,1) error:

$$\begin{aligned} X_t &= \mu_t + \sigma_t\,\epsilon_t, \quad \mathbb{E}[\epsilon_t] = 0,\, \mathbb{E}[\epsilon_t^2] = 1,\, \epsilon_t \text{ i.i.d.}, \\ \mu_t &= \lambda\, X_{t-1}, \\ \sigma_t^2 &= a_0 + a\,(X_{t-1} - \mu_{t-1})^2 + b\,\sigma_{t-1}^2. \end{aligned} \tag{2.3}$$

For stationarity we now have to add the condition $|\lambda| < 1$. As before, the innovations will be assumed to be either normally, Student-t_8 or Student-t_4 distributed, and we fix the parameters as $a_0 = 3 \cdot 10^{-6}$, $a = 0.05$ and $b = 0.92$ and moreover $\lambda = 0.04$, unless stated otherwise. We observe that the AR(1)-GARCH(1,1) process (2.3) can be rewritten in the form

$$\begin{aligned} X_t &= \sum_{k=0}^{\infty} \lambda^k\, Y_{t-k}, \\ Y_t &= \sigma_t\,\epsilon_t, \\ \sigma_t^2 &= a_0 + a\,Y_{t-1}^2 + b\,\sigma_{t-1}^2, \\ \mathbb{E}[\epsilon_t] &= 0,\, \mathbb{E}[\epsilon_t^2] = 1,\, \epsilon_t \text{ i.i.d.}, \end{aligned} \tag{2.4}$$

so that $(Y_t)_{t\in\mathbb{Z}}$ is a GARCH(1,1).

As a first step, we look at how the square-root-of-time rule performs for large data sets. We do this by simulating each of our models for a very long period, thereby obtaining good approximations for the scaled 1-day value-at-risk on the one hand, and the actual 10-day value-at-risk on the other. Unless stated otherwise, we will always consider the value-at-risk at the Basel confidence-level $\alpha = 99\%$. The results of these simulations are shown in Figure 2.

The top graph concerns the simple random walk models. For a random walk with normal innovations the $\sqrt{10}$ scaling rule is of course exact and the two empirically obtained values do indeed coincide in this case. For the heavier tailed Student-t_8 and Student-t_4 innovations, the square-root-of-time rule overestimates the true value-at-risk, with a bias which is greater the heavier the tail. This can at least qualitatively be understood from the Central Limit Theorem: for large N, Z_t^N will be approximately $\mathcal{N}(0, N\sigma_X^2)$ distributed, where σ_X^2 is the variance of $X_t \overset{d}{=} X$. We therefore expect that $\mathrm{VaR}_\alpha(Z^N) \simeq \sqrt{N}\sigma_X q_\alpha^{\mathcal{N}}$, where $q_\alpha^{\mathcal{N}}$ is the α quantile of the standard normal distribution. On the other hand, the scaled 1-day value-at-risk is $\sqrt{N}\sigma_X q_\alpha^X$, where q_α^X is the quantile of the normalised random variable $\sigma_X^{-1}X$. If the latter is heavy-tailed, then $q_\alpha^X > q_\alpha^{\mathcal{N}}$, and $\mathrm{VaR}_\alpha(Z^N) < \sqrt{N}\,\mathrm{VaR}_\alpha(X)$. Of course, this approximation will only be accurate for sufficiently large N, depending on α, and it is not a priori clear whether it will be relevant for $\alpha = 0.99$, $N = 10$; this particular point will be discussed in more detail in Section 5 below. More precise approximations in the random walk case can be found by using the Cornish-Fisher expansion, involving higher moments like skewness and kurtosis.

For the GARCH(1,1) process (2.2) (see second graph in Figure 2) we observe that for normally distributed innovations the square-root-of-time rule underestimates the 10-day

value-at-risk, while for the two Student ones the scaling rule leads to an overestimation. The latter phenomenon can be given a similar qualitative explanation as in the random walk case, but now based on the Martingale Central Limit Theorem: see Theorem 6.6 and Corollary 6.9 below. We see that the $\sqrt{10}$ scaling rule is in general not a perfect choice for *unlimited* data sets. This is in line with Christoffersen et al. [7] and Diebold et al. [8], who recommend for big data sets to directly model the 10-day returns instead of using scaling. However, we stress that from such investigations for big data sets we cannot draw recommendations for situations where the available data are limited to a few hundred observations, whence the present study.

Turning finally to the AR(1)-GARCH(1,1) model (2.3), the bottom graph of Figure 2 shows that for all three innovations under consideration, the (scaled) 1-day quantiles stay very close to the (scaled) 1-day quantiles for the corresponding GARCH(1,1) model shown in the graph above. This is explained by the fact that the unconditional (stationary) 1-day variances of the two processes will differ by a factor of only $(1-\lambda^2)^{-1}$ (cf. (6.9) below) and that therefore the 1-day value-at-risk for our AR(1)-GARCH(1,1) with $\lambda = 0.04$ will be a factor of only $(1-\lambda^2)^{-1/2} \simeq 1.0008$ times the GARCH(1,1) value. On the other hand, the actual 10-day value-at-risks for the AR(1)-GARCH(1,1) all show a slight increase by a factor of about $1+\lambda$, relative to the GARCH(1,1) case. This can, at least qualitatively, be explained by the fact that for small λ the 10-day variances of the two processes differ by a factor of $(1+\frac{9}{10}\lambda+O(\lambda^2))^2$, as will be seen in equation (6.10) below. In the case of normally distributed innovations, this increase of variance results in an underestimation, by the square-root-of-time rule, of the true 10-day value-at-risk, whereas for the (in practice more realistic) Student-t innovations the $\sqrt{10}$ rule works quite well (even extremely well in the case of the Student-t_8, although this is probably accidental, due to our particular choice of parameters). We see that for the Student-t_8 and Student-t_4 innovations, the increase in variance is apparently compensated by the fact that near the 1% quantile range the probability distribution function of the 10-day log-returns is 'lighter tailed' than that of the 1-day log-returns[4], in agreement with what the Martingale Central Limit Theorem would predict (cf. Section 6 below). Note that here we have an illustration of a phenomenon mentioned in the introduction, namely that of a non-zero auto-regressive parameter having a negligible effect on the 1-day value-at-risk, while having a sizeable effect on the 10-day one. Ironically, in the example considered here, this effect serves to rescue the $\sqrt{10}$-rule rather than to invalidate it.

3 Alternative empirical estimators

Our primary objective is to obtain good 10-day value-at-risk estimates from data sets of not more than 250 daily returns. It seems reasonable to explore potential alternatives to the square-root-of-time rule, and to investigate whether these, on average on data sets of this size, under- or outperform the simple scaling rule. In this section and the next we propose and investigate a number of natural alternative estimators, which are empirical in the sense that they will not explicitly use the structure of the data generating process (or, in a practical situation, the structure of a specific time-series model fitted to the

[4]However, farther out in the tails the 10-day returns will become much heavier tailed, according to Brummelhuis [5], Brummelhuis and Guégan [6].

daily data). By contrast, in Section 6, we will take the nature of the process as given, and look for theoretically motivated improvements on the $\sqrt{10}$-rule.

The different empirical value-at-risk estimators which will be studied are:

1. *Square-root-of-time rule.* Starting with the daily log-returns $X_1, \ldots, X_{250}$, we evaluate the empirical quantile. This 1-day value-at-risk is scaled with $\sqrt{10}$ to get an estimate for the 10-day value-at-risk.
2. *Non-overlapping periods.* Using the 25 non-overlapping 10-day log-returns: $Y^n := \sum_{i=0}^{9} X_{10n-i}$ $(n = 1, 2, \ldots, 25)$, the 99% quantile is estimated.
3. *Overlapping periods.* Using the 241 overlapping 10-day log-returns: $Y^n := \sum_{i=0}^{9} X_{n+i}$ $(n = 1, 2, \ldots, 241)$, the empirical quantile is evaluated.
4. *Random resampling.* Randomly, and with replacement, 10 out of the 250 daily log-returns $(X_{i_1}, \ldots, X_{i_{10}})$ are chosen, and the sum of these, $Y := \sum_{k=1}^{10} X_{i_k}$, is formed. This procedure is repeated 10 000 times, and the empirical quantile of these values $(Y^n)_{n=1,\ldots,10\,000}$ is evaluated. This is also known as *bootstrapping*.
5. *Independent resampling.* In this method, 10 weakly dependent values out of the 250 daily log-returns are chosen in the following way: pick $X_{i_1}, \ldots, X_{i_{10}}$ such that $\min_{j<k}(i_k - i_j) \geq 10$. The 10 chosen values are summed up to $Y := \sum_{k=1}^{10} X_{i_k}$, and the procedure is repeated 10 000 times. Finally the empirical quantile of the $(Y^n)_{n=1,\ldots,10\,000}$ is estimated.
6. *Dependent resampling.* Here, 10 strongly dependent values out of the 250 daily log-returns are chosen. For $n = 1, \ldots, 231$, pick $X^n_{i_1}, \ldots, X^n_{i_{10}}$, where $i_k \in [n, n + 19]$. Sum these 10 values to get $Y^n := \sum_{k=1}^{10} X^n_{i_k}$. In order to arrive at about 10 000 values as for methods (4) and (5) above, this procedure is repeated 44 times $(Y^{n,l}, l = 1, \ldots, 44)$ and the empirical quantile of the $(Y^{n,l})$ is evaluated.
7. *Extreme value method.* We use a generalised Pareto distribution (GPD) G_ξ, defined by
$$G_\xi(x) = \begin{cases} 1 - (1 + \xi\, x)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-x} & \text{if } \xi = 0, \end{cases}$$
with $x \geq 0$ if $\xi \geq 0$ and $0 \leq x \leq -1/\xi$ if $\xi < 0$, to fit the largest 10% and the smallest 10% of the data. For the remaining 80% of the distribution the empirical values are taken. From this new distribution, 10 values $(X_{i_1}, \ldots, X_{i_{10}})$ are simulated and summed up to $Y := \sum_{k=1}^{10} X^n_{i_k}$. This procedure is repeated 10 000 times and the empirical quantile of the Y's is evaluated.

Remark 3.1 By requiring $(i_k - i_j) \geq 10$ in estimation method 5, we expect that the dependence between the chosen values to be small: for example, in a second order stationary GARCH model, the conditional volatility forecast $\mathbb{E}[\sigma^2_{t+h} \mid \sigma^2_t] = a_0 \frac{1-(a+b)^h}{1-(a+b)} + (a+b)^h \sigma^2_t$ has an exponentially decreasing dependence on on the initial volatility σ^2_t. Similarly, it can be shown using (A.14) below that $\mathrm{Cov}(X^2_{t+h}, X^2_t)$ decreases exponentially as $h \to \infty$, and we take this as an indication of asymptotic independence. In fact, a stationary GARCH can be shown to be strongly mixing with geometric rate, cf. e.g. Basrak et al. [3], Theorem 3.1.

When comparing the performance of the estimators (1)-(7), we will work with $k = 250$ daily values. For analogous investigations for larger values of k it would be important to have consistent estimators, i.e. estimators converging to the true value-at-risk for

stationary time series, as the length of the time series increases. Since in our case the data set consists of a fixed number of $k = 250$ values, consistency is not the key issue. The aim is to minimise the deviation of an estimated value-at-risk from the true value-at-risk. Hence for small values of k the variance of an estimator will be equally important, as we will see below. The main issue here is the sample size effect.

For each estimator we evaluate the mean difference between the estimated quantile and the true 10-day value-at-risk, $\mathrm{VaR}^{10}_{99\%}$, which in this section we will simply denote by $\mathrm{VaR}_{99\%}$. The quotient of this difference and the corresponding difference involving the square-root-of-time rule will measure the relative performance of the chosen alternative estimator. More precisely, for each of the stochastic processes under consideration we simulate 1000 independent paths of length 250. If for the i-th simulation the square-root-of-time estimator of $\mathrm{VaR}_{99\%}$ is denoted by x^i, and the alternative estimators described above are denoted by y^i_j $(j = 2, \ldots, 7)$ we evaluate S_j defined by:

$$S_j := \frac{\sum_{i=1}^{1000} |y^i_j - \mathrm{VaR}_{99\%}|}{\sum_{i=1}^{1000} |x^i - \mathrm{VaR}_{99\%}|};$$

S_j is a measure for the relative goodness of the j-th estimator relative to the first one. If $S_j > 1$, then the mean of the absolute difference between the alternative estimator and the true value-at-risk is bigger than the corresponding difference for the $\sqrt{10}$-rule. In such a situation, the $\sqrt{10}$-rule should clearly be preferred. More generally, $S_j > S_k$ means that the k-th method outperforms the j-th method, relative to the square-root-of-time rule.

In addition, we compute some descriptive statistics for each of the estimators individually, including $j = 1$ (note that $x^i = y^i_1$). The first one is the mean relative deviation of the j-th value-at-risk estimator from the true value-at-risk

$$S^0_j = \frac{1}{1000} \sum_{i=1}^{1000} \frac{|y^i_j - \mathrm{VaR}_{99\%}|}{\mathrm{VaR}_{99\%}};$$

note that $S_j = S^0_j / S^0_1$. We also evaluate the (signed) difference between the mean of the estimators $(y^i_j)_{i=1,\ldots,1000}$ and the true value-at-risk $\mathrm{VaR}_{99\%}$:

$$S^1_j = \frac{\overline{y}_j - \mathrm{VaR}_{99\%}}{\mathrm{VaR}_{99\%}},$$

where $\overline{y}_j = \frac{1}{1000} \sum_{i=1}^{1000} y^i_j$. Next, S^2_j will measure the relative mean squared error of estimator number (j):

$$S^2_j = \frac{1}{1000} \sum_{i=1}^{1000} \Big(\frac{y^i_j - \mathrm{VaR}_{99\%}}{\mathrm{VaR}_{99\%}} \Big)^2,$$

while S^3_j is a skewness-type quantity:

$$S^3_j = \frac{1}{1000} \sum_{i=1}^{1000} \Big(\frac{y^i_j - \mathrm{VaR}_{99\%}}{\mathrm{VaR}_{99\%}} \Big)^3,$$

Finally, we introduce the kurtosis-related measure

$$S^4_j = \frac{1}{1000} \sum_{i=1}^{1000} \Big(\frac{y^i_j - \mathrm{VaR}_{99\%}}{\mathrm{VaR}_{99\%}} \Big)^4.$$

These numbers will be helpful to understand the statistical properties of the estimation methods listed above.

4 Performance of the estimators

In this section we compare the performance of the square-root-of-time rule with the alternative value-at-risk estimators introduced above, on the particular processes introduced in Section 2. We start with an example. Figure 3 shows the comparison between the square-root-of-time rule and the method using overlapping periods (method 3) for a random walk with normal innovations. Every point in the graph represents one of the thousand simulations we did. The x-values represent the $\mathrm{VaR}_{99\%}$ estimates for the square-root-of-time rule, while the y-axis shows the corresponding risk estimates based on overlapping 10-day periods. The horizontal and the vertical line mark the true value-at-risk $\mathrm{VaR}_{99\%}$ for the normal random walk. The dashed line has slope 1 and goes through $(\mathrm{VaR}_{99\%}, \mathrm{VaR}_{99\%})$. The continuous line goes through $(\mathrm{VaR}_{99\%}, \mathrm{VaR}_{99\%})$ as well, but with slope $S_j = \frac{S_j^0}{S_1^0} = \frac{\sum_{i=1}^{1000} |y_j^i - \mathrm{VaR}_{99\%}|}{\sum_{i=1}^{1000} |x^i - \mathrm{VaR}_{99\%}|}$ (in the case of Figure 3, $j = 3$, i.e. the overlapping periods method). S_j is the ratio of the mean deviations of the value-at-risk estimates (x^i) and (y_j^i) from the true value-at-risk, as introduced in Section 3. In Figure 3 we observe that the slope of the continuous line is clearly bigger than 1, and in fact close to 2. This means that the mean difference of a 10-day value-at-risk estimate using overlapping 10-day log-returns is about twice the difference we get when applying the square-root-of-time rule. If neither of the two methods would outperform the other, the continuous line would coincide with the dashed line.

Each of the Figures 4–12 displays for a fixed underlying process the comparison of the square-root-of-time rule with the alternative quantile estimators. We observe that for the methods working directly with the available 10-day log-returns (methods 2 and 3; left-hand graphs in each of the Figures 4 to 12) the deviation of y-values is much larger than the deviation of x-values, which results in slopes $S_j \gg 1$. This indicates that for all processes under investigation, scaling with $\sqrt{10}$ is better than directly using the actually occurred 10-day returns. The heavier tailed the distribution of the random term of the process is, the more the slope of the continuous line for these two alternative estimators decreases. But also for these heavier tailed cases, scaling with $\sqrt{10}$ is still better than using directly 10-day intervals. The clearness of this result is striking, especially for the GARCH(1,1) and AR(1)-GARCH(1,1) models, which contain a certain dependence structure that is neglected when applying the square-root-of-time scaling rule. Hence one might have expected that the methods working with real 10-day log-returns and thus taking dependence into consideration would outperform the $\sqrt{10}$ scaling rule. But apparently, the uncertainty of estimating a one-in-a-hundred event based on only a few values has a huge unfavourable impact on the estimator's performance. Additional investigations showed that the poor performance of methods 2 and 3 cannot be improved by replacing empirical quantiles by GPD estimates.

Also method 6 (dependent resampling, top right graphs), which aims at keeping some dependence, has a relatively poor performance. Further investigations made clear that this cannot be improved considerably by increasing the number of repetitions which was 44 in our simulations (leading to about 10 000 values). The results for the remaining

three methods (4, 5 and 7, i.e. random resampling, independent resampling and extreme value method) do in general not differ much from each other. Exceptions are the models with Student-t_4 innovations (see Figures 6, 9 and 12), where the extreme value method seriously underestimates the tail index $\alpha(=1/\xi)$ in some of the simulations, which results in a slight increase of the slope for the EVT method, compared to those for random resampling and independent resampling.

These estimators 4, 5 and 7 perform particularly well for the heavy-tailed independent models (random walks with Student-t innovations). Method 4 (random resampling) is the easiest of the three estimators to implement, and it is not outperformed by either of the two others (independent resampling and extreme value method). Hence we concentrate on comparing the square-root-of-time rule with random resampling (top middle graphs). For all three types of models (random walk, GARCH(1,1), AR(1)-GARCH(1,1)) with normal innovations, the two methods perform equally well. When log-returns are heavier tailed (in our case Student-t_8 and Student-t_4), random resampling performs much better than the $\sqrt{10}$ rule in the random walk model, slightly better for the GARCH(1,1) processes, and about equally well for the AR(1)-GARCH(1,1) processes. This is due to the fact that for random walks, working with the sum of 10 independent values as done in the random resampling gives an unbiased estimator, whereas scaling a one-day value-at-risk of a heavy tailed distribution by $\sqrt{10}$ leads to an overestimation of the 10-day value-at-risk. This overestimation is much smaller for GARCH(1,1) and AR(1)-GARCH(1,1) processes, as we also already saw in Figure 2.

Figure 13 summarises the information contained in the slopes in Figures 4 to 12. This graphical representation of the values S_j^0 shows that for all nine processes under investigation, the square-root-of-time rule performs at least reasonably well, even though it does not always reach the performance of the random resampling method. As expected, for models with light tails none of the other methods outperforms the square-root-of-time scaling.

The various graphs in Figures 4 to 12 actually contain more information than just the slope. For most methods, the majority of points lies below the horizontal line drawn at $\mathrm{VaR}_{99\%}$. To investigate this visual impression, we compute the descriptive statistics S_j^i introduced before. In Figures 14 to 17 the corresponding values are shown in dotplots.

We start by analysing the bias S_j^1. In Figure 14 an unbiased estimation method would produce a point on the dotted zero line. We can observe that methods 2–7 tend to underestimate value-at-risk (negative values of S_j^1), whereas the square-root-of-time rule is the only one tending to overestimate value-at-risk. For random walks, the bias for the square-root-of-time rule gets worse for heavier tailed innovations. This behavior stays in line with the following consideration. Since the distribution of the sum of N independent Student-t distributed random variables tends to a normal distribution for increasing N, the quantiles of the 10-day log-return can be approximated by those of a normal distribution: $\mathrm{VaR}_\alpha^{\text{true,10-day}} \approx \mathrm{VaR}_\alpha^{\mathcal{N},\text{10-day}} = \sqrt{10}\,\mathrm{VaR}_\alpha^{\mathcal{N},\text{1-day}}$. Hence the relative error S_1^1 made when scaling the 99% quantile of a Student-t distribution with ν degrees of freedom can be approximated by $\frac{\sqrt{10}\,\mathrm{VaR}_{99\%}^{t_\nu,\text{1-day}} - \sqrt{10}\,\mathrm{VaR}_{99\%}^{\mathcal{N},\text{1-day}}}{\sqrt{10}\,\mathrm{VaR}_{99\%}^{\mathcal{N},\text{1-day}}}$, which equals $\frac{q_{99\%}^{t_\nu} - q_{99\%}^{\mathcal{N}}}{q_{99\%}^{\mathcal{N}}}$. For $\nu = 8$ and $\nu = 4$ this gives the theoretical values 0.078 and 0.139, respectively, which are quite close to the values 0.06 and 0.10 observed in the simulations (see Figure 14).

In Figure 15 we see the reason for the fact that random resampling and independent resampling provide good estimates of value-at-risk (see Figures 4 to 12). These two methods have lowest variabilities. For random and independent resampling also the third and the fourth moment of the difference from the true value are relatively low (see Figures 16 and 17). Low third and fourth moments imply that these estimators are stable. Note that the main issue is not consistency, but variability. In a practical situation there is only one single (small) data set available. Hence an estimator method with a low variability should be preferred to a consistent one with a huge variability (if there is a large bias, this should of course be taken into account). This implies that measures similar to the mean squared error (MSE) are the right ones to use in this context. As a clarifying example, we compare the results for dependent resampling and independent resampling. In Figure 14 we observe that dependent resampling gives for GARCH-type processes *in the mean* a better result than independent resampling, but its variability with respect to the true value-at-risk $\mathrm{VaR}_{99\%}$ over a size 1000 sample is bigger than that of the independent resampling (see Figure 15). In a practical situation we would prefer the latter – as can be seen in Figures 4 to 12, where independent resampling always yields a smaller slope S than dependent resampling.

One should be aware of another peculiarity when analysing the measures S_j^i. We highlight this by means of the GARCH(1,1) model. For dependent resampling (method 6) the absolute values of S_6^1 are quite small (see Figure 14). For this method the mean of the estimates is close to the true value-at-risk. Since the distribution of the estimates is skewed (see Figures 7 to 9), the median of these estimates is much smaller than the true value. This explains why the slope S (which depends strongly on the median of the estimates) for this method is so large (see again Figures 7 to 9), while the value of S_6^1 is quite small. Reasoned by the fact that the AR(1)-GARCH(1,1) process in equation (2.3) can be rewritten in the form (2.4), most peculiarities when estimating value-at-risk in GARCH(1,1) models also hold for AR(1)-GARCH(1,1) models. In particular, the descriptive statistics S_j^i for the two processes differ only slightly.

The comments we made based on the slope S (see Figure 13) stay in line with the conclusions which can be drawn from the descriptive statistics S_1^i (Figures 14 to 17). Overall we can conclude that for estimating unconditional 10-day value-at-risk numbers at a level 99%, the $\sqrt{10}$ rule works quite well. Random resampling is as good, and sometimes slightly better.

5 The importance of the confidence level α

The relatively good performance of the square-root-of-time rule reported on in the previous section came as something of a surprise, in the light of work previous work of McNeil and Frey [13] and of Brummelhuis and Guégan [6] on conditional value-at-risk. The latter, in particular, computed the asymptotic behavior of conditional N-day VaR_α in the case of a normal GARCH(1,1) as one lets the confidence parameter α tend to 1, and found that in this limit, $\mathrm{VaR}_\alpha(Z^N) \simeq c_N \cdot (\mathrm{VaR}_\alpha)^N$, for any fixed N, with explicit constant $c_N > 0$. This clearly implies that the square-root-of-time rule fails dramatically in such a limit. However, if $N = 10$, these asymptotics are apparently not relevant for a fixed α like 0.99, apart from the other obvious difference with [6], [13], namely that

in the present paper we study the unconditional value-at-risk instead of the conditional one.

5.1 When is the CLT relevant and when are tail asymptotics?

To investigate this point in more detail, we turn to the simpler situation of a non-normal random walk. From the Central Limit Theorem we know that the normalised sum of n independent and identically distributed random variables with finite variance converges weakly to a standard normal distribution as n tends to infinity. In practical applications, one would typically like to approximate the sum of n independent and identically distributed random variables by a normal distribution, if n is reasonably large. Here, the immediately question arises of what n is large enough if, for example, we are interested in evaluating the α-th quantile for a given $\alpha \in (0,1)$?

To find an answer to this question, we numerically study the convolution of Student-t distributed random variables. Let $X_1, \ldots, X_n$ denote independent copies of a Student-t distributed random variable with ν degrees of freedom, expectation 0 and variance 1. Let $S := \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$ denote the standardised sum, and F_S the corresponding cumulative distribution function. We compare the quantiles $s_\alpha := F_S^{-1}(\alpha)$ of the sum with the quantiles $q_\alpha := \Phi^{-1}(\alpha)$ of a standard normal distribution. We first do this for $\nu = 8$ degrees of freedom. The contour plot in Figure 18 shows the area, where q_α is a good approximation of s_α. The x-values represent the number of convolutions n (on a logarithmic scale). On the y-axis $1-\alpha$ can be read off. The range of values for the level α goes from 0.50 (top) to $1 - 10^{-7}$ (bottom). The lines (in pairs) show the range, where the approximation error $\epsilon := |\log \frac{s_\alpha}{q_\alpha}|$ equals a certain threshold. For example for the sum of $n = 8$ Student-t_8 distributed random variables, the only levels for which a normal distribution yields a very good approximation ($\epsilon \leq 0.01$) are the ones with $\alpha \in [0.897, 0.984]$ (and for symmetry reasons also $\alpha \in [0.016, 0.103]$). Allowing for an error $\epsilon \leq 0.05$, for $n = 8$ all quantiles with $\alpha \in [0.0008, 0.9992]$ can be approximated by normal quantiles. In order to read off the quantiles easier for small values of n, we plot the same lines a second time, using a linear scale for the α-values, see Figure 19. For the original Student-t_8 distribution ($n = 1$), asking for an error of at most $\epsilon = 0.01$, we observe that only quantiles in the range $\alpha \in [0.959, 0.971]$ (and $\alpha \in [0.029, 0.041]$) can be replaced by normal quantiles. For all levels between 0.041 and 0.959, normal quantiles exceed Student-t_8 quantiles, while for $\alpha > 0.971$ (and $\alpha < 0.029$) replacing Student-t_8 quantiles by normal ones leads to an underestimation (in absolute values).

Repeating this comparison for a Student-t distribution with $\nu = 4$ degrees of freedom yields the expected outcome. The sum must be taken over a bigger sample (n large) in order that quantiles can be closely approximated by normal ones, see Figures 20 and 21. For example for $n = 16$ terms, allowing for an error $\epsilon \leq 0.01$, the range of good approximations only contains the levels $\alpha \in [0.957, 0.980]$ (and $\alpha \in [0.020, 0.043]$).

On the other hand, going far into the tails of the distribution of the convolutions, the power behaviour always persists. For independent random variables $(X_i)_{i=1,\ldots,n}$ with $P[X_i > x] \sim c\,x^{-\nu}$ as $x \to \infty$ (as is the case for Student-t_ν distributed random variables), the same power decay holds for the standardised sum $S = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$: $P[S > s] \sim n\ c(\sqrt{n}s)^{-\nu}$ as $s \to \infty$. Hence high quantiles $s_\alpha = F_S^{-1}(\alpha)$ of the standardised

sum can be approximated by the quantiles $x_\alpha := F_{t_\nu}^{-1}(\alpha)$ of the original distribution: $s_\alpha \sim n^{1/\nu-1/2}\, x_\alpha$ for n large, as $\alpha \to 1$.

Similar to the above investigation for approximations with a normal distribution (Figures 18–21), the lower limits of α can be calculated, from where on the approximation error $\epsilon := |\log \frac{s_\alpha}{n^{1/\nu-1/2}\, x_\alpha}|$ is small. For Student-t distributions with $\nu = 8$ and $\nu = 4$ degrees of freedom, the corresponding lines for $\epsilon \in \{0.1, 0.2, 0.5, 1.0\}$ are drawn in Figures 22 and 23. For $\nu = 4$ we observe that already for the sum of two random variables ($n = 2$) one has to go very far into the tail ($\alpha \in [0.997, 1]$ and $\alpha \in [0, 0.003]$) in order to get an approximation error ϵ smaller than 0.1. For $\nu = 8$, the power decay of the tail has an even weaker influence for the convolutions. Here, for $n = 2$, the approximation error is smaller than 0.1 only for levels $\alpha \in [0.99996, 1]$ (and $\alpha \in [0, 0.00004]$).

These investigations make clear that the results for quantile estimation strongly depend on the choice of the level α. In particular, analysing a 99% quantile yields very different results from investigating the asymptotic behaviour as α tends to one.

5.2 Scaling a 1-day 95% VaR to a 10-day 99% VaR.

Another problem which is related to the confidence level α is the question of how to scale a 1-day 95% value-at-risk to a 10-day 99% value-at-risk. In absence of further knowledge on the 1-day return distributions, a straightforward methodology would be to multiply the 1-day 95% VaR with the quotient $q^{\mathcal{N}}_{99\%}/q^{\mathcal{N}}_{95\%}$ (where $q^{\mathcal{N}}_\alpha$ denotes the α quantile of a standard normal distribution), and then scale the resulting value with the square-root-of time. But this first step – multiplying with the quotient – is in general not appropriate, as a short investigation shows.

Taking a random walk with Student-t_4 innovations as an illustrative example, we can read off from Figure 21 that using the quantile of a normal distribution as an approximation for the true α quantile yields an underestimation of about 13% for $\alpha = 99\%$, and an overestimation of more than 8% for $\alpha = 95\%$. Hence the error committed when multiplying with $q^{\mathcal{N}}_{99\%}/q^{\mathcal{N}}_{95\%}$ is more than 20%. For 10-day quantiles, the corresponding error is about 7%, composed of an underestimation of 5% and an overestimation of 2%. One could now argue that proceeding the other way around – first scaling the 1-day 95% VaR with the square-root-of-time and then committing an error of (only) 7% by multiplying with $q^{\mathcal{N}}_{99\%}/q^{\mathcal{N}}_{95\%}$ – leads to a less serious error. But the problem here is that already the first step – multiplying a 1-day 95% VaR with the square-root-of-time – can produce a rather big estimation error for realistic models with dependent log-returns. While for $\alpha = 99\%$ the overestimation of quantiles caused by the square-root-of-time rule is partially compensated for by the dependence in the model (which increases the 10-day quantiles), at the $\alpha = 95\%$-level the square-root-of-time rule causes an underestimation of the quantiles, which gets even worse for dependent log-returns.

These considerations make clear that transforming a 1-day 95% value-at-risk into a 10-day 99% value-at-risk is rather delicate. One should first transform the 1-day 95% estimate appropriately into a 1-day 99% estimate, before applying the square-root-of-time scaling rule. An appropriate transformation from one quantile level to the other one requires knowledge of the tail of the one-day distribution, which corresponds to the recommendation to start directly with a 99% quantile level for daily log-returns.

6 Multiple-period quantile estimation in GARCH(1,1) and AR(1)-GARCH(1,1) models

In this section we try to obtain some theoretical insight into the scaling behaviour of the 99% value-at-risk in GARCH(1,1) and AR(1)-GARCH(1,1) models with the dual aim of, on the one hand, better understanding the relatively good performance of $\sqrt{10}$-rule, and, on the other hand, finding simple algorithms which might improve upon this rule and serve as alternatives to performing extensive Monte Carlo simulations. Before embarking upon this, however, we first raise the opposite question: *in which kind of a situation will the square-root-of-time be less satisfactory, and even downright misleading?* There are at least two potential reasons for such a bad performance: the presence of trends or of auto-regression in the returns. Either of these, even when negligible on a 1-day scale, can lead to considerable bias of the $\sqrt{10}$-rule on the 10-day scale.

We begin with trends. Throughout this paper we assume a zero trend in our models, but it is not difficult to account for a non-zero one. First, one estimates the trend over a 1-day horizon:

$$\widehat{\mu} := \frac{1}{n}\sum_{t=1}^{n} X_t.$$

Next, one estimates the 10-day value-at-risk for the centered log-returns $\widetilde{X}_t := X_t - \widehat{\mu}$. Calling the latter $\mathrm{VaR}_\alpha(\widetilde{X})$, the 10-day value-at-risk of the original time series can be estimated by simply subtracting the 10-day trend:

$$\mathrm{VaR}_\alpha(X) = \mathrm{VaR}_\alpha(\widetilde{X}) - 10\widehat{\mu}.$$

The following example shows that trends can have an important effect over 10-days, even when over a single day they are quite negligible:

Example 6.1 Consider independent 1-day log-returns $(X_t)_{t\in\mathbb{Z}}$ with $\mathcal{N}(0.1\%, (1\%)^2)$ distribution. The method described above leads to the correct 10-day 99% value-at-risk of 6.36%. Direct multiplication of the 1-day value-at-risk with $\sqrt{10}$ gives the wrong value 7.04%, a relative error of approximately 10%. More generally, in the normal frame-work the error $\mathrm{VaR}_\alpha^{10} - \sqrt{10}\,\mathrm{VaR}_\alpha^1$ from using the $\sqrt{10}$-rule without accounting for trends is $(\sqrt{10} - 10)\mu$. □

Turning to the effect of auto-regression, it can be argued that one reason for the relatively good performance of the square-root-of-time rule in a (trend-less) GARCH(1,1) model is the fact that successive log-returns are uncorrelated. The interdependence between returns is only indirect, through the variances and squared returns, which influences the absolute size of the successive log-returns, but not their signs, which remain completely random. In an AR(1)-GARCH(1,1) model, this changes. In such a model, the direct linear dependence between successive values causes the $\sqrt{10}$-rule to be increasingly off as the AR(1)-parameter λ increases. The following example illustrates this for a classical AR(1)-model with i.i.d. normal errors, in the absence of any heteroskedasticity. We will consider the full AR(1)-GARCH(1,1) later.

Example 6.2 Consider the following 1-day log-return process $(X_t)_{t\in\mathbb{Z}}$:

$$X_t = \frac{1}{2}X_{t-1} + \epsilon_t,$$

where $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Then $X_t = \sum_{s=0}^{\infty} 2^{-s}\epsilon_{t-s}$, and the X_t's are therefore normally distributed: $X_t \sim \mathcal{N}(0, \frac{4}{3})$. Applying the square-root-of-time rule leads to a 10-day 99% value-at-risk estimate of 8.495. To calculate the correct value-at-risk, we have to analyze the 10-day log-returns $Z_t^{10} = \sum_{k=0}^{9} X_{t-k}$, which are easily seen to be normally distributed as well, with $Z_t^{10} \sim \mathcal{N}(0, 34.672)$ (cf. the proof of Lemma 6.3 below). Hence the true 10-day 99% value-at-risk equals 13.698, which is much larger than the $\sqrt{10}$-estimate. □

This example easily generalises:

Lemma 6.3 *Consider the following stationary AR(1) for the daily log-returns* $(X_t)_{t\in\mathbb{Z}}$*:*

$$X_t = \lambda X_{t-1} + \epsilon_t, \quad \textit{where } |\lambda| < 1 \textit{ and } \epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

Let $Z_t^{10} = \sum_{k=0}^{9} X_{t-k}$ *be the 10-day log-returns . Then the 10-day value-at-risk* $VaR_{99\%}^{10} := VaR_{99\%}(Z_t^{10})$ *equals*

$$VaR_{99\%}^{10} = \frac{1}{1-\lambda}\sqrt{10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}}\, q_{99\%}^{\mathcal{N}}, \tag{6.1}$$

where $q_{99\%}^{\mathcal{N}}$ *denotes the 99% quantile of the standard normal distribution.*

Proof: Since $X_t = \sum_{s=0}^{\infty} \lambda^s \epsilon_{t-s}$, we have that $X_t \sim \mathcal{N}\big(0, \frac{1}{1-\lambda^2}\big)$. The 10-day log-return can be computed as

$$Z_t^{10} = \sum_{k=0}^{9}\sum_{s=0}^{\infty} \lambda^s \epsilon_{t-s-k} = \sum_{k=0}^{8} \frac{1-\lambda^{k+1}}{1-\lambda}\epsilon_{t-k} + \frac{1-\lambda^{10}}{1-\lambda}\sum_{k=9}^{\infty}\lambda^{k-9}\epsilon_{t-k}.$$

Since the ϵ_t's are independent and normally distributed, this leads to

$$Z_t^{10} \sim \mathcal{N}\left(0, \frac{1}{(1-\lambda)^2}\left(10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}\right)\right),$$

which corresponds to a 10-day 99% value-at-risk of

$$\mathrm{VaR}_{99\%}^{10} = \frac{1}{1-\lambda}\sqrt{10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}}\, q_{99\%}^{\mathcal{N}}.$$

□

Applying the square-root-of-time rule leads to a 10-day 99% value-at-risk estimate of

$$\widehat{\mathrm{VaR}}_{99\%}^{10} = \sqrt{\frac{10}{1-\lambda^2}}\, q_{99\%}^{\mathcal{N}}. \tag{6.2}$$

It can be shown that (6.2) underestimates the true 10-day value-at-risk given in (6.1) by a relative error of

$$\frac{\mathrm{VaR}_{99\%}^{10} - \widehat{\mathrm{VaR}}_{99\%}^{10}}{\mathrm{VaR}_{99\%}^{10}} > 0.68\lambda.$$

When λ is large, we therefore cannot expect the square-root-of-time rule to perform well. Fortunately, in practice, when fitting AR(1)-GARCH(1,1) models, one typically gets a low value of λ. The largest one we found when fitting daily log-returns for foreign exchange rates, stock indices, single stocks and 10-year government bonds was an estimate of $\lambda = 0.14$. For most data sets, the coefficient λ takes values around 0.04.

Returning to GARCH processes, we now re-examine the issue of computing the N-day 99% quantile from first principles. As a first step, we compute the conditional and unconditional variances of the N-day log-returns in stationary GARCH(1,1) and AR(1)-GARCH(1,1) models. Recall that the GARCH(1,1) (2.2) is second-order stationary if and only if $a + b < 1$. For an AR(1)-GARCH(1,1) like (2.3), we also need $|\lambda| < 1$. To simplify the formulas below, it is convenient to introduce the parameter[5]

$$c := a + b. \tag{6.3}$$

We put

$$\sigma_\infty^2 := \frac{a_0}{1-c} = \frac{a_0}{1-a-b}, \tag{6.4}$$

the variance of a stationary GARCH.

Lemma 6.4 *For the GARCH(1,1) process $(X_t)_{t\in\mathbb{Z}}$ defined by in (2.2) with $c < 1$, the conditional variance of the N-day returns is equal to*

$$\Sigma_N^2 := Var(\sum_{t=1}^N X_t \mid \sigma_1) = N\sigma_\infty^2 + \frac{1-c^N}{1-c}(\sigma_1^2 - \sigma_\infty^2). \tag{6.5}$$

The unconditional (stationary) N-day variance is

$$Var(\sum_{t=1}^N X_t) = N\sigma_\infty^2. \tag{6.6}$$

The analogous result for AR(1)-GARCH(1,1) processes is:

Lemma 6.5 *For the AR(1)-GARCH(1,1) process $(X_t)_{t\in\mathbb{Z}}$ defined by (2.3) with $c < 1$ and $\lambda \neq c \neq \lambda^2$, the conditional N-day variance is given by:*

$$\begin{aligned}\widetilde{\Sigma}_N^2 &:= Var(\sum_{t=1}^N X_t \mid \mu_1, \sigma_1) \\ &= \frac{1}{(1-\lambda)^2}\Big\{\sigma_\infty^2\Big(N - 2\lambda\frac{\lambda^N-1}{\lambda-1} + \lambda^2\frac{\lambda^{2N}-1}{\lambda^2-1}\Big) + \\ &\qquad +(\sigma_1^2-\sigma_\infty^2)\Big(\frac{1-c^N}{1-c} - 2\lambda\frac{\lambda^N-c^N}{\lambda-c} + \lambda^2\frac{\lambda^{2N}-c^N}{\lambda^2-c}\Big)\Big\}.\end{aligned} \tag{6.7}$$

If $\lambda = c$ or $\lambda^2 = c$, the variance can be found by taking the appropriate limit in (6.7). Finally, the unconditional N-day variance is

$$\sigma_{(N)}^2 := Var(\sum_{t=1}^N X_t) = \frac{\sigma_\infty^2}{(1-\lambda)^2}\Big(N - 2\lambda\frac{1-\lambda^N}{1-\lambda^2}\Big). \tag{6.8}$$

These lemmas are probably well-known, but we nevertheless included proofs in Appendix A, for convenience of the reader. For $N = 1$ and $N = 10$, (6.8) leads to

$$\sigma_{(1)} = \sigma_\infty\sqrt{\frac{1}{1-\lambda^2}} \simeq (1 + \frac{1}{2}\lambda^2)\sigma_\infty, \tag{6.9}$$

[5]Note that c is precisely the auto-regressive parameter when rewriting X_t^2 in the form of ARMA(1,1): $X_t^2 = a_0 + (a+b)X_{t-1}^2 - b\eta_{t-1} + \eta_t$, with $\eta_t := X_t^2 - \sigma_t^2$ a martingale.

and

$$\sigma_{(10)} = \frac{\sigma_\infty}{1-\lambda}\sqrt{10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}} \simeq \sigma_\infty\sqrt{10}\,(1+\frac{9}{10}\lambda), \tag{6.10}$$

assuming λ to be close to 0. These formulas were used in the discussion of Figure 2 at the end of Section 2.

In order to convert variance or volatility into value-at-risk, we have to know the relevant quantiles of the underlying standardised probability distribution. These are hard to obtain theoretically for the N-day returns in a GARCH(1,1) or an AR(1)-GARCH(1,1) model. However, for a fixed confidence level α and sufficiently large N, we can gain some insight from the Martingale Central Limit Theorem. For completeness, we state and prove the following theorem in both the conditional and unconditional case, although we are principally interested in the latter.

Theorem 6.6 *Let $(X_t)_{t\in\mathbb{Z}}$ a GARCH(1,1) process with $\mathbb{E}[\epsilon^4] < \infty$, such that $c < 1$ and $c_\epsilon := a^2\mathbb{E}[\epsilon^4] + 2ab + b^2 < 1$. Then we have that, conditional on σ_1 and in the sense of weak convergence,*

$$\frac{1}{\Sigma_N}\sum_{t=1}^{N} X_t \mid \sigma_1 \xrightarrow{d} \mathcal{N}(0,1) \quad (N\to\infty), \tag{6.11}$$

and

$$\frac{1}{\sqrt{N}\sigma_\infty}\sum_{t=1}^{N} X_t \mid \sigma_1 \xrightarrow{d} \mathcal{N}(0,1) \quad (N\to\infty). \tag{6.12}$$

We also have, unconditionally, that

$$\frac{1}{\sqrt{N}\sigma_\infty}\sum_{t=1}^{N} X_t \xrightarrow{d} \mathcal{N}(0,1) \quad (N\to\infty). \tag{6.13}$$

Proof: The proof uses a form of the Martingale Central Limit Theorem due to McLeish [12]; see Appendix A for details.

Remark 6.7 For Student-t_ν distributed innovations ϵ with variance 1, $\mathbb{E}[\epsilon^4]$ can be computed as:

$$\mathbb{E}[\epsilon^4] = \begin{cases} \frac{3\nu-6}{\nu-4} & \text{if } \nu > 4, \\ \infty & \text{if } \nu \le 4. \end{cases}$$

For normally distributed ϵ the fourth moment is 3.

Theorem 6.6 generalises to AR(1)-GARCH(1,1) processes:

Theorem 6.8 *For $(X_t)_{t\in\mathbb{Z}}$ an AR(1)-GARCH(1,1) with $|\lambda| < 1$ and parameters a and b satisfying the same conditions as in Theorem 6.6 we have, conditional on μ_1 and σ_1, that*

$$\frac{1}{\widetilde{\Sigma}_N}\sum_{t=1}^{N} X_t \mid \mu_1, \sigma_1 \xrightarrow{d} \mathcal{N}(0,1) \quad (N\to\infty), \tag{6.14}$$

and

$$\frac{1}{\sigma_{(N)}}\sum_{t=1}^{N} X_t \mid \mu_1, \sigma_1 \xrightarrow{d} \mathcal{N}(0,1) \quad (N\to\infty), \tag{6.15}$$

while also

$$\frac{1}{\sigma_{(N)}} \sum_{t=1}^{N} X_t \xrightarrow{d} \mathcal{N}(0,1) \quad (N \to \infty), \tag{6.16}$$

unconditionally.

Proof: By reduction to the previous theorem: see Appendix A. □

It is known that the operation of taking quantiles is sequentially continuous with respect to weak convergence of probability distributions, provided that the limiting distribution is continuous: cf. Resnick [14] or Embrechts et. al. [10], Proposition A1.7. We therefore conclude from Theorem 6.8:

Corollary 6.9 *For an AR(1)-GARCH(1,1) process, under the conditions of Theorem 6.8 and for any fixed α, the unconditional N-day value-at-risk VaR_α^N satisfies*

$$\frac{VaR_\alpha^N}{\sigma_{(N)}} \to q_\alpha^{\mathcal{N}} \quad (N \to \infty), \tag{6.17}$$

where $q_\alpha^{\mathcal{N}}$ is, as before, the α-th quantile of the standard normal distribution, and where $\sigma_{(N)}$ is given by (6.8)*. A similar result holds for the conditional N-day value-at-risk:*

$$\frac{VaR_{c,\alpha}^N}{\widetilde{\Sigma}_N} \quad (N \to \infty). \tag{6.18}$$

Similar results hold for a GARCH(1,1), by taking $\lambda = 0$.

On the basis of these results, we expect that for N large the unconditional N-day value-at-risk can be approximated using a normal distribution:

$$\mathrm{VaR}_{99\%}^N \approx \sigma_{(N)}\, q_{99\%}^{\mathcal{N}}.$$

In particular, for $N = 10$:

$$\mathrm{VaR}_{99\%}^{10} \approx \frac{\sigma_\infty}{1-\lambda}\sqrt{10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}}\; q_{99\%}^{\mathcal{N}}. \tag{6.19}$$

Remark 6.10 Note that the denominators in (6.17) and (6.18) can be replaced by $\sqrt{N}\sigma_\infty/(1-\lambda)$, which is aymptotically equivalent to both $\sigma_{(N)}$ and $\widetilde{\Sigma}_N$ as $N \to \infty$. However, for finite (but sufficiently large) N it is probably better to keep $\widetilde{\Sigma}_N$ respectively $\sigma_{(N)}$, depending on whether one is interested in conditional or unconditional value-at-risk. Also, for conditional value-at-risk one might want to correct for the $\sum_{t=1}^N X_t$ having a non-zero (conditional) mean equal to $\mu_1(1-\lambda^N)/(1-\lambda)$, by formula (A.5) below. Since we are primarily interested in unconditional value-at-risk, we will not pursue this further.

In the case of a GARCH(1,1), the normal approximation of Corollary 6.9 introduces the desired scaling of $\sqrt{N}$ for big N, since $\sigma_{(N)} = \sqrt{N}\sigma_\infty$ if $\lambda = 0$. It does not, however, validate the $\sqrt{N}$-rule, since the 1-day unconditional value-at-risk will be given by $\mathrm{VaR}_{99\%}^1 = \sigma_\infty q_{99\%}^{\epsilon}$, and the 99% quantile of the innovations $\epsilon_t \overset{d}{=} \epsilon$ will be bigger than that of the normal quantile if ϵ is heavy-tailed. The conclusion is that the $\sqrt{N}$-rule will significantly over-estimate risk if N is big.

We are not, however, primarily interested in the validity of this rule for arbitrarily large values of N, but rather in its performance for $N = 10$, and since 10 is still a fairly small number, we cannot necessarily expect the normal approximation (6.19) to perform well. To investigate this further, we numerically compare (6.19) with the true 10-day value-at-risk for our familiar list of test processes: random walks, GARCH(1,1)and AR(1)-GARCH(1,1) processes, with both normal and Student-t innovations. The unconditional 10-day value-at-risk can be decomposed as

$$\mathrm{VaR}^{10}_{99\%} = \sigma_{(10)}\, q^{10}_{99\%},$$

where $\sigma_{(10)}$ is the standard deviation of the stationary distribution, given by (6.10), and where $q^{10}_{99\%}$ denotes the quantile of the *standardised* stationary 10-day distribution. For each of the models under consideration, we simulate 10 000 000 periods of 10 days, to get good approximations for the corresponding value-at-risk numbers. We divide them by the relevant standard deviations, which are $\sqrt{10}$ for the random walk, $\sqrt{10}\sigma_\infty$ for the GARCH(1,1) process, and $\sqrt{10 - 2\lambda\frac{1-\lambda^{10}}{1-\lambda^2}}\,\frac{\sigma_\infty}{1-\lambda}$ in the AR(1)-GARCH(1,1) case. In Figure 24, the circles designate the relative errors of the standardised 10-day 99% quantile with the 99% normal one. We first observe that the normal approximation systematically underestimates true value-at-risk, with a relative error which is less than 5% for all of the random walks considered, and also for the two GARCH(1,1) and AR(1)-GARCH(1,1) processes with normal innovations. For Student-t innovations, the relative error can double in size, going up to 10% for a t_4. Obviously, for the latter, $N = 10$ is not yet sufficiently large for N-day log-returns to be close to a normal distribution in the 99% quantile range.

We have also compared the 99% quantile of the standardised stationary 10-day distributions with those of the standardised stationary 1-day distributions, denoted by q^1_α. The crosses in Figure 24 show the relative difference $(q^{10}_{99\%} - q^1_{99\%})/q^{10}_{99\%}$ between these standardised 10-day and 1-day quantiles. We observe that for the GARCH(1,1) and AR(1)-GARCH(1,1) processes (although not for the simple random walks), this leads to a smaller relative error than simply using the normal quantile. Hence, the $\sqrt{10}$-rule still wins from the normal approximation of Corollary 6.9 with $N = 10$.

Clearly, the distribution of the normalised sums $\sigma^{-1}_{(10)}\sum_{k=1}^{10} X_k$ still differs appreciably from a standard normal one in the 1% quantile range. In order to improve the approximation, the shape of the 10-day distribution should be taken into consideration. Because of symmetry reasons, all odd moments are zero in our models. Hence once the variance is known, the fourth moment gives the next most important shape-information. We therefore compute the kurtosis of the N-day stationary log-returns, to begin with for a GARCH(1,1). Recall that the kurtosis of a mean-zero random variable X is defined as $\kappa_X := \mathbb{E}[X^4]/\mathbb{E}[X^2]^2$. Also recall that $c = a + b$ and $c_\epsilon = a\mathbb{E}(\epsilon^4) + 2ab + b^2$.

Lemma 6.11 *For a stationary GARCH(1,1) model $(X_t)_{t\in\mathbb{Z}}$ such that $c, c_\epsilon < 1$, the kurtosis of $Z_N := \sum_{k=1}^N X_k$ is equal to:*

$$\kappa_N = 3\,\frac{N-1}{N} + \frac{\mathbb{E}[\epsilon^4]}{N}\,\frac{1-c^2}{1-c_\epsilon} + \frac{6}{N^2}\,\frac{N - \frac{1-c^N}{1-c}}{1-c}\,\Big((a\,\mathbb{E}[\epsilon^4] + b)\frac{1-c^2}{1-c_\epsilon} - c\Big).$$

Proof: See (A.18) and (A.19) in the appendix. □

A similar result holds for AR(1)-GARCH(1,1) processes. Since in practice the autoregressive parameter is small, we will only compute the first order contribution in λ:

Lemma 6.12 *For a stationary AR(1)-GARCH(1,1) process* $(X_t)_{t\in\mathbb{Z}}$ *such that* $|\lambda|, c, c_\epsilon < 1$, *the kurtosis of* $Z_N := \sum_{k=1}^N X_k$ *is equal to:*

$$\begin{aligned}\kappa_N = 3\frac{N-1}{N} + \frac{\mathbb{E}[\epsilon^4]}{N}\frac{1-c^2}{1-c_\epsilon} + \\ + 6\frac{N+4\lambda}{N^3}\left((a\,\mathbb{E}[\epsilon^4]+b)\,\frac{1-c^2}{1-c_\epsilon} - c\right)\frac{N - \frac{1-c^N}{1-c} - 2\lambda(1-c^{N-1})}{1-c} + \\ + O(\lambda^2).\end{aligned}$$

Proof: See (A.22) and (A.23) in the appendix. □

To obtain an improved approximation for the 10-day value-at-risk, we look for a ν such that the corresponding Student-t_ν distribution has the same kurtosis as Z_{10}. Given $\kappa_N > 3$, $\nu \equiv \nu(N)$ can be found explicitly as $\nu = \frac{4\kappa_N - 6}{\kappa_N - 3}$. We then expect $\sigma_{(10)}\, q_{99\%}^{t_\nu}$ to be a good approximation for the true 99% quantile of Z_{10}. Hence, for a model with known parameters a_0, a, b and (small) λ, and known $\mathbb{E}[\epsilon^4]$, we obtain a relatively explicit and rapidly computable approximation which can be used as an alternative to both the square-root-of-time rule and a full-scale Monte-Carlo simulation. It can also be used to assess the effect on value-at-risk due to the uncertainty inherent in any statistical estimation procedure for the parameters, by letting the latter vary over their statistical 95% confidence intervals; doing so with Monte-Carlo would be much more time-consuming.

We note in passing that we could of course also have used other parametric families of distributions, like the Pearson family, to fit the first four moments. An advantage of using t_ν is that it has the expected heavy tails, and that an accurate approximation for its quantiles is explicitly known: if $\nu \geq 4$, $q_{99\%}^{t_\nu}$ can be approximated by $2.324 + 6.44/\nu + 5.27/\nu^2 + 65.7/\nu^3$, with a relative error less than 0.2% for each such ν.

We turn to a numerical test of the new approximation. In Figures 25 (simulated values) and 26 (simulated versus approximated values; using Lemmas 6.11 and 6.12) the 10-day value-at-risk in AR(1)-GARCH(1,1) models is shown for various choices of λ (keeping $a = 0.05$ and $b = 0.92$ fixed). A first observation is that simulated and approximated values almost coincide. Another observation – very relevant for practical applications – is the fact that for realistic (small) values of λ the square-root-of-time scaling rule still yields a very close approximation (white symbols) of the true 10-day value-at-risk (black symbols)[6]. For larger values of λ, the $\sqrt{10}$-rule is increasingly unreliable, while the kurtosis-based approximation still stays close to the true value-at-risk.

[6]In particular, for Student-t_8 innovations, $\lambda = 0.04$, $a = 0.05$ and $b = 0.92$, the square-root-of-time approximation coincides with the true 10-day value-at-risk, an observation already made at the end of Section 2. The improved approximation based on Lemma 6.12 provides a reasonably good quantitative explanation of this.

7 Conclusion

As we just noted, although not perfect, the $\sqrt{10}$-rule still performs remarkably well as long as λ is not too big. To either confirm or invalidate this observation, we have further compared $\sqrt{10}\,\mathrm{VaR}^1_{99\%}$ graphically with the true $\mathrm{VaR}^{10}_{99\%}$ for GARCH parameters a and b ranging over the intervals $[0, 0.0625]$ and $[0, 0.94]$ (keeping the other parameter fixed at, respectively, 0.92 and 0.05). As we see from Figures 27 and 28 (varying a) and Figures 29 and 30 (varying b), the square-root-of-time rule still yields a decent approximation over these quite wide ranges of parameters, and this the better the smaller a and the bigger b is.

The, perhaps somewhat surprising, conclusion of this paper is that, in the context of AR(1)-GARCH(1,1) return processes, and as long as auto-regressive effects in the returns, or their squares, are not too big (meaning a small λ and a), the $\sqrt{10}$-rule does a good job of predicting unconditional 10-day value-at-risk at the 99% confidence level, both when compared with other possible empirical estimators on relatively small data sets of size 250, and theoretically when taken the nature of the process as given. Slight improvements can be made by using bootstrapping techniques when estimating empirically, or by using an improved approximation based on the 10-day kurtosis, but on the whole scaling by $\sqrt{10}$ performs remarkably well. This conclusion is somewhat at variance with that of Christoffersen et al. [7] and Diebold et al. [8], which were also based on studies of a GARCH model. However, these papers were principally concerned with the behavior of *conditional volatility*, while we are interested in the *unconditional quantile*, at a specific confidence-level and for a specific time-window.

It remains to be seen whether the conclusions of this paper remain valid for other return processes than the ones considered here. We refer to Kaufmann [11], Chapter 3, for a study of the $\sqrt{10}$-rule, on its own and when compared with the other empirical estimators from Section 3, for the class of stochastic volatility models, with and without jumps. There seems to have been very little work done on theoretical multiple period value-at-risk estimation in such models. We also note that the investigation from Section 4 can, in principle, be repeated for any sufficiently long empirical return series without imposing any time-series model (except of course for a necessary hypothesis of stationarity). For the moment, however, we can summarize the findings of this paper as follows, paraphrasing Diebold et al. [8]: *Converting 1-day value-at-risk to N-day value-at-risk: scaling by $\sqrt{N}$ may be worse than you think, but is not as bad as you might fear.*

References

[1] Alexander C. (1998) Volatility and Correlation: Measurement, Models and Applications. In: *Risk Management and Analysis*, Vol. 1 (ed. Alexander C.), Wileys, 125–171.

[2] Basel Committee on Banking Supervision (1996) *Overview of the amendment to the capital accord to incorporate market risk.*

[3] Basrak, B, Davis, R. A. and Mikosch, T. (2002) *Regular Variation of GARCH Processes*, Stoch. Proc. Appl. 99 (2002), 95–115.

[4] Bollerslev T. (1986) *Generalized Autoregressive Conditional Heteroscedasticity*, Journal of Econometrics, 31, 307–327.

[5] Brummelhuis R.G.M. (2004) *Inter-temporal Dependence in ARCH(1)-models*, working paper, Birkbeck College.

[6] Brummelhuis R.G.M. and Guégan D. (2000) *Multi-period Conditional Distribution Functions for Heteroscedastic Models with Applications to VaR*, prépublication 00.13, Université de Reims, Dépt. de math, submitted.

[7] Christoffersen P.F., Diebold F.X. and Schuermann T. (1998) *Horizon Problems and Extreme Events in Financial Risk Management*, Federal Reserve Bank of New York, 109–118.

[8] Diebold F.X., Hickman A., Inoue A. and Schuermann T. (1997) *Converting 1-day Volatility to h-day Volatility: Scaling by $\sqrt{h}$ is Worse than You Think*, Wharton Financial Institutions Center, Working Paper 97–34, http://fic.wharton.upenn.edu/fic/

[9] Drost F.C. and Nijman T.E. (1993) *Temporal Aggregation of GARCH Processes*, Econometrica, 61, 909–927.

[10] Embrechts P., Klüppelberg C. and Mikosch T. (1997) *Modelling Extremal Events for Insurance and Finance*, Springer, New York.

[11] Kaufmann, R. (2004) *Long-Term Risk Management*, PhD Thesis, ETH Zürich.

[12] McLeish D.L. (1974) *Dependent Central Limit Theorems and Invariance Principles*, The Annals of Probability, Vol. 2, No. 4, 620–628.

[13] McNeil A.J. and 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.

[14] Resnick S.I. (1987) *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

[15] Stărică C. (2003) *Is Garch(1,1) as Good a Model as the Accolades of the Nobel Prize Would Imply?*, Preprint, Chalmers University of Technology, Gothenburg.

Appendix A Proofs

We start with the conditional variance of an GARCH(1,1) or AR(1)-GARCH(1,1) process. Since $\mathbb{E}[\sigma_t^2 \mid \sigma_{t-s}] = \mathbb{E}[a_0 + a\,(\sigma_{t-1}\,\epsilon_{t-1})^2 + b\,\sigma_{t-1}^2 \mid \sigma_{t-s}] = a_0 + c\,\mathbb{E}[\sigma_{t-1}^2 \mid \sigma_{t-s}]$, where $c = a + b$, recursion gives the well-known formula

$$\mathbb{E}[\sigma_t^2 \mid \sigma_{t-s}] = a_0 \frac{1-c^s}{1-c} + c^s\,\sigma_{t-s}^2 \qquad \text{for } s \geq 0. \tag{A.1}$$

It follows that the unconditional stationary variance is given by

$$\mathbb{E}[\sigma_t^2] = \frac{a_0}{1-c} = \sigma_\infty^2. \tag{A.2}$$

□

Proof of Lemma 6.4: The conditional N-day variance is given by

$$\begin{aligned}\Sigma_N^2 :&= \mathrm{Var}(\sum_{t=1}^N X_t \mid \sigma_1) = \mathbb{E}[(\sum_{t=1}^N X_t)^2 \mid \sigma_1] - (\mathbb{E}[\sum_{t=1}^N X_t \mid \sigma_1])^2 \\ &= \sum_{t=1}^N \mathbb{E}[X_t^2 \mid \sigma_1].\end{aligned}$$

The individual terms are known from (A.1):

$$\mathbb{E}[X_t^2 \mid \sigma_1] = \mathbb{E}[\sigma_t^2 \mid \sigma_1] = a_0 \frac{1-c^{t-1}}{1-c} + c^{t-1}\,\sigma_1^2 = \sigma_\infty^2 + c^{t-1}(\sigma_1^2 - \sigma_\infty^2).$$

Hence

$$\Sigma_N^2 = N\,\sigma_\infty^2 + \sum_{t=1}^N c^{t-1}(\sigma_1^2 - \sigma_\infty^2) = N\,\sigma_\infty^2 + (\sigma_1^2 - \sigma_\infty^2)\frac{1-c^N}{1-c}, \tag{A.3}$$

proving (6.5). The unconditional stationary N-day variance can then be computed as

$$\mathrm{Var}(\sum_{t=1}^N X_t) = \mathbb{E}[\mathrm{Var}(\sum_{t=1}^N X_t \mid \sigma_1)] + \mathrm{Var}(\mathbb{E}[\sum_{t=1}^N X_t \mid \sigma_1]) = N\,\sigma_\infty^2.$$

□

Remark A.1 We observe in passing that σ_1^2 influences the conditional variance of the N-day return much more strongly than it does the expected variance of the day N-return: if, for example, $1-c = 1/N$ (which, for $N = 10$, is not an unreasonable value, empirically), and if we approximate $(1-N^{-1})^N \simeq e^{-1}$, then (A.3) gives $\Sigma_N \simeq N(1-e^{-1})\sigma_1^2 + e^{-1}N\sigma_\infty^2$, while σ_1^2 enters only with a factor of $(1-N^{-1})^N \simeq e^{-1}$ into $\mathbb{E}[X_N^2|\sigma_1] = \mathbb{E}[\sigma_N^2|\sigma_1]$; if c is such that $(1-c^N)/(1-c) \simeq N$, we even get $\Sigma_N^2 \simeq N\sigma_1^2$.

Proof of Lemma 6.5: As before,

$$\widetilde{\Sigma}_N^2 = \mathbb{E}[(\sum_{k=1}^N X_k)^2 \mid \mu_1, \sigma_1] - (\mathbb{E}[\sum_{k=1}^N X_k \mid \mu_1, \sigma_1])^2. \tag{A.4}$$

Using recursively $X_t = \mu_t + \sigma_t\,\epsilon_t$ and $\mu_t = \lambda\,X_{t-1}$, we get

$$\begin{aligned}\mathbb{E}[\sum_{k=1}^{N} X_k \mid \mu_1, \sigma_1] &= \mu_1 \sum_{k=1}^{N} \lambda^{k-1} + \mathbb{E}[\sum_{k=1}^{N} \sigma_k \epsilon_k \sum_{s=0}^{N-k} \lambda^s \mid \mu_1, \sigma_1] \\ &= \mu_1 \frac{1-\lambda^N}{1-\lambda}.\end{aligned} \tag{A.5}$$

The first conditional expectation in (A.4) can be rewritten as

$$\begin{aligned}\mathbb{E}[(\sum_{k=1}^{N} X_k)^2 \mid \mu_1, \sigma_1] &= \mathbb{E}[(\mu_1 \sum_{k=1}^{N} \lambda^{k-1} + \sum_{k=1}^{N} \sigma_k \epsilon_k \sum_{s=0}^{N-k} \lambda^s)^2 \mid \mu_1, \sigma_1] \\ &= \Big(\mu_1 \frac{1-\lambda^N}{1-\lambda}\Big)^2 + \sum_{k=1}^{N} \Big(\frac{1-\lambda^{N-k+1}}{1-\lambda}\Big)^2 \mathbb{E}[\sigma_k^2 \mid \sigma_1],\end{aligned}$$

and hence

$$\widetilde{\Sigma}_N^2 = \sum_{k=1}^{N} \Big(\frac{1-\lambda^{N-k+1}}{1-\lambda}\Big)^2 \mathbb{E}[\sigma_k^2 \mid \sigma_1].$$

Plugging in formula (A.1) for the conditional variance yields[7]

$$\widetilde{\Sigma}_N^2 = \sum_{k=1}^{N} \Big(\frac{1-\lambda^{N-k+1}}{1-\lambda}\Big)^2 \Big(a_0 \frac{1-c^{k-1}}{1-c} + c^{k-1}\,\sigma_1^2\Big).$$

Since $\sigma_\infty^2 = a_0/(1-c)$, we find

$$\widetilde{\Sigma}_N^2 = \frac{\sigma_\infty^2}{(1-\lambda)^2} \sum_{k=1}^{N} (1-\lambda^{N-k+1})^2 + \frac{\sigma_1^2 - \sigma_\infty^2}{(1-\lambda)^2} \sum_{k=1}^{N} (1-\lambda^{N-k+1})^2 c^{k-1}.$$

If $|\lambda| < 1$, terms of the form $\sum_{k=1}^{N} (1-\lambda^{N-k+1})^2$ can be written as

$$\sum_{k=1}^{N} (1-\lambda^{N-k+1})^2 = \sum_{k=1}^{N} (1 - 2\lambda^{N-k+1} + \lambda^{2N-2k+2}) = N - 2\lambda \frac{1-\lambda^N}{1-\lambda} + \lambda^2 \frac{1-\lambda^{2N}}{1-\lambda^2},$$

while if $\lambda \neq c$, $\lambda^2 \neq c$,

$$\begin{aligned}\sum_{k=1}^{N} (1-\lambda^{N-k+1})^2 c^{k-1} &= \sum_{k=1}^{N} (1 - 2\lambda^{N-k+1} + \lambda^{2N-2k+2}) c^{k-1} \\ &= \frac{1-c^N}{1-c} - 2\lambda \frac{\lambda^N - c^N}{\lambda - c} + \lambda^2 \frac{\lambda^{2N} - c^N}{\lambda^2 - c}.\end{aligned}$$

Hence for $|\lambda| < 1$, $\lambda \neq c$, $\lambda^2 \neq c$

$$\begin{aligned}\widetilde{\Sigma}_N^2 = \frac{1}{(1-\lambda)^2} \Bigg\{ &\sigma_\infty^2 \Big(N - 2\lambda \frac{1-\lambda^N}{1-\lambda} + \lambda^2 \frac{1-\lambda^{2N}}{1-\lambda^2}\Big) + \\ &+ (\sigma_1^2 - \sigma_\infty^2) \Big(\frac{1-c^N}{1-c} - 2\lambda \frac{\lambda^N - c^N}{\lambda - c} + \lambda^2 \frac{\lambda^{2N} - c^N}{\lambda^2 - c}\Big) \Bigg\}.\end{aligned} \tag{A.6}$$

(Observe, as a check, that this reduces to (A.3), if $\lambda = 0$.)

[7] Note that $\sigma_t^2 = a_0 + (a\epsilon_t^2 + b)\sigma_{t-1}^2$, as for a GARCH(1,1).

Similarly, for $\lambda = c$ we get

$$\widetilde{\Sigma}_N^2 = \frac{1}{(1-\lambda)^2}\Big\{\sigma_\infty^2\Big(N - 2\lambda\frac{1-\lambda^N}{1-\lambda} + \lambda^2\frac{1-\lambda^{2N}}{1-\lambda^2}\Big) + \\ +(\sigma_1^2 - \sigma_\infty^2)\Big(\frac{(1+\lambda^{N+1})(1-\lambda^N)}{1-\lambda} - 2N\lambda^N\Big)\Big\},$$

while if $\lambda^2 = c$:

$$\widetilde{\Sigma}_N^2 = \frac{1}{(1-\lambda)^2}\Big\{\sigma_\infty^2\Big(N - 2\lambda\frac{1-\lambda^N}{1-\lambda} + \lambda^2\frac{1-\lambda^{2N}}{1-\lambda^2}\Big) + \\ +(\sigma_1^2 - \sigma_\infty^2)\Big(\frac{1-\lambda^{2N}}{1-\lambda^2} - 2\lambda^N\frac{1-\lambda^N}{1-\lambda} + N\lambda^{2N}\Big)\Big\}.$$

For the unconditional N-day variance we obtain for $|\lambda| < 1$

$$\begin{aligned}\mathrm{Var}(\sum_{t=1}^N X_t) &= \mathbb{E}[\mathrm{Var}(\sum_{t=1}^N X_t \mid \mu_1, \sigma_1)] + \mathrm{Var}(\mathbb{E}[\sum_{t=1}^N X_t \mid \mu_1, \sigma_1]) \\ &= \mathbb{E}[\widetilde{\Sigma}_N^2] + \mathrm{Var}\Big(\frac{1-\lambda^N}{1-\lambda}\mu_1\Big) \\ &= \frac{\sigma_\infty^2}{(1-\lambda)^2}\Big(N - 2\lambda\frac{1-\lambda^N}{1-\lambda} + \lambda^2\frac{1-\lambda^{2N}}{1-\lambda^2}\Big) + \Big(\frac{1-\lambda^N}{1-\lambda}\Big)^2\mathbb{E}[\mu_1^2],\end{aligned}$$

since $\mathbb{E}[\mu_1] = 0$. Using that

$$\begin{aligned}\mathbb{E}[\mu_1^2] &= \lambda^2\mathbb{E}[(\sum_{k=0}^\infty \lambda^k Y_{-k})^2] = \lambda^2\mathbb{E}[\sum_{k=0}^\infty (\lambda^k\,\sigma_{-k})^2] = \lambda^2\sum_{k=0}^\infty \lambda^{2k}\,\mathbb{E}[\sigma_{-k}^2] \\ &\overset{(A.2)}{=} \lambda^2\sum_{k=0}^\infty (\lambda^2)^k\,\sigma_\infty^2 = \sigma_\infty^2\,\frac{\lambda^2}{1-\lambda^2},\end{aligned}$$

this finally yields

$$\sigma_{(N)}^2 := \mathrm{Var}(\sum_{t=1}^N X_t) = \frac{\sigma_\infty^2}{(1-\lambda)^2}(N - 2\lambda\frac{1-\lambda^N}{1-\lambda^2}). \tag{A.7}$$

□

Proof of Theorem 6.6:

This will be based on a result from McLeish [12], which we state for the reader's convenience.

$$(Y_{N,j})_{N=1,2,\ldots,\ 1\le j\le k_N}$$

be a triangular array of random variables on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}_j \subset \mathcal{F}_{j+1}$ be an increasing filtration of sub-σ-algebras of $\mathcal{F}$, such that each $Y_{N,j}$ is $\mathcal{F}_j$-measurable[8].

[8]McLeish [12] allows the filtrations to depend on N, but that generality will not be needed here.

Theorem A.2 *(cf. McLeish* [12]*, Corollary 2.13) Suppose that*

(i) $(Y_{N,j})_{j,N}$ *is a* martingale difference array*, meaning that* $\mathbb{E}[Y_{j,N}|\mathcal{F}_{j-1}] = 0$*;*
(ii) $(Y_{N,j})_{j,N}$ *satisfies the Lindeberg condition: for each* $\varepsilon > 0$*,*

$$\lim_{N\to\infty} \sum_{j=1}^{k_N} \mathbb{E}[Y_{N,j}^2 \mathbf{1}_{\{|Y_{N,j}|>\varepsilon\}}] = 0; \tag{A.8}$$

(iii) If $\sigma_{N,j}^2 := \mathbb{E}[Y_{N,j}^2]$*, then*

$$\lim_{N\to\infty} \sum_{j=1}^{k_N} \sigma_{N,j}^2 = 1; \tag{A.9}$$

(iv)

$$\limsup_{N\to\infty} \sum_{1\le j\ne k\le k_N} \mathbb{E}[Y_{N,j}^2 Y_{N,k}^2] \le 1. \tag{A.10}$$

Then $\sum_{j=1}^{k_N} Y_{j,N}$ *converges weakly to* $\mathcal{N}(0,1)$ *as* $N \to \infty$*.*

Remark A.3 In McLeish [12], this is proved as a corollary of a much more general Central Limit Theorem for martingales, establishing convergence in distribution to the normal under much weaker conditions. However, Theorem A.2 seems particularly well adapted for applications to econometric models, especially when bounds on variances and higher moments and cross-moments can be readily derived. The present case of a GARCH(1,1) is a typical example.

Proof of Theorem 6.6. We will first prove (6.11), (6.12). Given our GARCH(1,1) process $(X_t)_{t\in\mathbb{Z}}$, we define a triangular array by

$$Y_{N,j} := \Sigma_N^{-1} X_j,\ 1 \le j \le N,$$

where Σ_N is the conditional variance computed in (6.11). (We do not assume the GARCH(1,1) process to be stationary already.) We will verify the hypotheses of McLeish's Theorem A.2. First of all, we clearly have a martingale difference array. Moreover, $\sum_{j=1}^{N} \mathbb{E}[X_j^2|\sigma_1] = \Sigma_N$, so (A.9) is trivially verified. The Lindeberg condition (A.8) will, as usual, be a consequence of a uniform bound on some higher order moment, for which it is convenient to take the fourth order one. Assuming that $\mathbb{E}[\epsilon^4] := \mathbb{E}[\epsilon_k^4] < \infty$, put

$$c_\epsilon := a^2\mathbb{E}[\epsilon^4] + 2ab + b^2, \tag{A.11}$$

and, assuming that $c_\epsilon < 1$, define the constant

$$\mu_{4,\infty} := \frac{a_0^2 + 2a_0 c\sigma_\infty^2}{1-c_\epsilon}\,\mathbb{E}[\epsilon^4] = \sigma_\infty^4\,\frac{1-c^2}{1-c_\epsilon}\,\mathbb{E}[\epsilon^4]; \tag{A.12}$$

$\mu_{4,\infty}^2$ is simply the fourth moment of the stationary GARCH(1,1), as can be seen from the following lemma:

Lemma A.4 *Let* $(X_k)_{k\ge1}$ *be a GARCH(1,1) such that both* $c < 1$ *and* $c_\epsilon < 1$*. Then the conditional fourth moment of* X_k *is*

$$\mathbb{E}[X_k^4|\sigma_1] = \mu_{4,\infty} + c_\epsilon^{k-1}(\sigma_1^4\mathbb{E}[\epsilon^4] - \mu_{4,\infty}) + 2a_0 c\mathbb{E}[\epsilon^4]\frac{c_\epsilon^{k-1} - c^{k-1}}{c_\epsilon - c}(\sigma_1^2 - \sigma_\infty^2), \tag{A.13}$$

while the unconditional one is simply $\mu_{4,\infty}$*.*

Bollerslev [4], Theorem 2, gives a recursive procedure to compute all unconditional moments.

We see that, in particular, the fourth moments can be bounded independently of k, which is all we need for (A.8): writing $\mathbb{E}_{\sigma_1}$ for $\mathbb{E}[\cdot|\sigma_1]$, we estimate in the usual way: $\sum_{j\leq N}\mathbb{E}_{\sigma_1}\left[Y_{N,j}^2\mathbf{1}_{\{|Y_{j,N}|>\varepsilon\}}\right]\leq\varepsilon^2\sum_{j\leq N}\mathbb{E}_{\sigma_1}[Y_{N,j}^4]\leq \text{Const}\cdot\varepsilon^2 N/\Sigma_N^4\to 0$, $N\to\infty$.

Proof of Lemma A.4: Starting off with the variance process $\sigma_k^2=a_0+(a\mathbb{E}[\epsilon^2]+b)\sigma_{k-1}^2$ we find, after squaring

$$\mathbb{E}[\sigma_k^4|\sigma_1]=a_0^2+2a_0c\mathbb{E}[\sigma_{k-1}^2|\sigma_1]+c_\epsilon\mathbb{E}[\sigma_{k-1}^4|\sigma_1].$$

Iterating, and using (A.1), we get

$$\begin{aligned}\mathbb{E}[\sigma_k^4|\sigma_1] &= a_0^2\left(\frac{1-c_\epsilon^{k-1}}{1-c_\epsilon}\right)+2a_0c\sum_{j=0}^{k-2}c_\epsilon^j\,\mathbb{E}[\sigma_{k-j-1}^2|\sigma_1]+c_\epsilon^{k-1}\sigma_1^4\\ &= (a_0^2+2a_0c\sigma_\infty^2)\frac{1-c_\epsilon^{k-1}}{1-c_\epsilon}+2a_0c\frac{c_\epsilon^{k-1}-c^{k-1}}{c_\epsilon-c}(\sigma_1^2-\sigma_\infty^2)+c_\epsilon^{k-1}\sigma_1^4,\end{aligned}$$

and the lemma follows using $X_k=\sigma_k\epsilon_k$ and conditional independence of ϵ_k. □

We finally check (A.10). For this, we will use the following lemma. Recall that $c=a+b$.

Lemma A.5 *For a GARCH(1,1) process $(X_n)_n$ with $\mathbb{E}[\epsilon^4]<\infty$ we have for all $j\geq 1$, k,*

$$\mathbb{E}[X_k^2X_{k-j}^2]=a_0\frac{1-c^j}{1-c}\mathbb{E}[X_{k-j}^2]+(a+\frac{b}{\mathbb{E}[\epsilon^4]})c^{j-1}\mathbb{E}[X_{k-j}^4],\tag{A.14}$$

and the same result with $\mathbb{E}[\cdot|\sigma_1]$ instead of $\mathbb{E}$, provided of course that $k-j\geq 1$.

Proof of Lemma A.5: By induction on j. First, if $j=1$, then

$$\mathbb{E}[X_k^2X_{k-1}^2]=\mathbb{E}[(a_0+aX_{k-1}^2+b\sigma_{k-1}^2)\epsilon_k^2X_{k-1}^2].$$

Using $\mathbb{E}[\sigma_{k-1}^2\epsilon_k^2X_{k-1}^2]=\mathbb{E}[\sigma_{k-1}^2\epsilon_k^2\sigma_{k-1}^2\epsilon_{k-1}^2]=\mathbb{E}[\sigma_{k-1}^4]=\mathbb{E}[X_{k-1}^4]/\mathbb{E}[\epsilon^4]$ leads to

$$\mathbb{E}[X_k^2X_{k-1}^2]=a_0\,\mathbb{E}[X_{k-1}^2]+(a+\frac{b}{\mathbb{E}[\epsilon^4]})\,\mathbb{E}[X_{k-1}^4],$$

for all k, as required. Next, suppose that $j\geq 2$, and that (A.14) holds for $j-1$ and arbitrary k. Then:

$$\mathbb{E}[X_k^2X_{k-j}^2]=\mathbb{E}[(a_0+aX_{k-1}^2+b\sigma_{k-1}^2)\epsilon_k^2X_{k-j}^2]=a_0\,\mathbb{E}[X_{k-j}^2]+c\,\mathbb{E}[X_{k-1}^2X_{k-j}^2],$$

since $\mathbb{E}[\sigma_{k-1}^2\,X_{k-j}^2]=\mathbb{E}[X_{k-1}^2\,X_{k-j}^2]$, because of the independence of ϵ_{k-1}^2 and $\sigma_{k-1}^2\,X_{k-j}^2$ for $j\geq 2$, and $\mathbb{E}[\epsilon_{k-1}^2]=1$. Plugging in (A.14) for $k-1$, $j-1$ then gives

$$\begin{aligned}\mathbb{E}[X_k^2X_{k-j}^2] &= a_0\,\mathbb{E}[X_{k-j}^2]+c\left(a_0\frac{1-c^{j-1}}{1-c}\mathbb{E}[X_{k-j}^2]+(a+\frac{b}{\mathbb{E}[\epsilon^4]})c^{j-2}\mathbb{E}[X_{k-j}^4]\right)\\ &= a_0\frac{1-c^j}{1-c}\mathbb{E}[X_{k-j}^2]+(a+\frac{b}{\mathbb{E}[\epsilon^4]})c^{j-1}\mathbb{E}[X_{k-j}^4].\end{aligned}$$

□

Turning to (A.10), and simply writing again $\mathbb{E}_{\sigma_1}$ for $\mathbb{E}[\cdot|\sigma_1]$, we compute

$$\begin{aligned}
&\sum_{1\le j\neq k\le k_N} \mathbb{E}_{\sigma_1}[Y_{N,j}^2 Y_{N,k}^2|\sigma_1] \\
&\quad = \frac{2}{\Sigma_N^4} \sum_{1\le j<k\le N} \mathbb{E}_{\sigma_1}[X_j^2 X_k^2|\sigma_1] \\
&\quad = \frac{2}{\Sigma_N^4} \sum_{j=1}^{N-1}\sum_{k=j+1}^{N} \left\{\sigma_\infty^2(1-c^{k-j})\mathbb{E}_{\sigma_1}[X_j^2] + c^{k-j-1}(a+\mathbb{E}[\epsilon^4]^{-1}b)\mathbb{E}_{\sigma_1}[X_j]^4\right\} \\
&\quad = \frac{2}{\Sigma_N^4} \sum_{j=1}^{N-1} \Bigg\{(N-j)\sigma_\infty^2\mathbb{E}_{\sigma_1}[X_j^2] - c\sigma_\infty^2\frac{1-c^{N-j}}{1-c}\mathbb{E}_{\sigma_1}[X_j^2] + \\
&\qquad\qquad + (a+\mathbb{E}[\epsilon^4]^{-1}b)\frac{1-c^{N-j}}{1-c}\mathbb{E}_{\sigma_1}[X_j^4]\Bigg\},
\end{aligned}$$

where we used (A.14) and $\sigma_\infty^2 = a_0/(1-c)$. Now by Lemma A.4, the fourth moments of the GARCH(1,1) are uniformly bounded, provided $c, c_\epsilon < 1$, and the same is of course true for $\mathbb{E}_{\sigma_1}[X_j^2|\sigma_1] = \sigma_\infty^2 + c^{j-1}(\sigma_1^2 - \sigma_\infty^2)$. It follows that

$$\begin{aligned}
\sum_{1\le j\neq k\le k_N} \mathbb{E}_{\sigma_1}[Y_{N,j}^2 Y_{N,k}^2|\sigma_1] &= \frac{2}{\Sigma_N^4}\sum_{j=1}^{N-1}\left\{(N-j)\sigma_\infty^2\left(\sigma_\infty^2 + c^{j-1}(\sigma_1^2-\sigma_\infty^2)\right) + O(N)\right\} \\
&= \frac{N(N-1)\sigma_\infty^4}{\Sigma_N^4} + O\left(\frac{1}{N}\right) \to 1,
\end{aligned}$$

as $N\to\infty$, since $\Sigma_N \simeq \sigma_\infty\sqrt{N}$ for large N, by Lemma 6.4. This establishes (A.10) and thereby (6.11). Finally, (6.12) is a direct consequence of (6.11) and of $\frac{\Sigma_N}{\sqrt{N}\sigma_\infty} \to 1 \quad (N\to\infty)$. □

Proof of (6.13): Similar, but (computationally) simpler, since we are now dealing with unconditional moments throughout. □

Proof of Theorem 6.8:

Observe that $(X_n)_n$ is no longer a martingale. However, the process $(X_n - \lambda X_{n-1})_n$ is a GARCH(1,1), to which we can apply Theorem 6.6. Hence,

$$\frac{1}{\Sigma_N}\sum_{j=1}^{N}(X_j - \lambda X_{j-1}) \xrightarrow{d} \mathcal{N}(0,1).$$

The sum can be re-arranged as

$$\frac{1-\lambda}{\Sigma_N}\sum_{j=1}^{N-1} X_j + \frac{-\lambda X_0 + X_N}{\Sigma_N}.$$

Since, clearly, $X_N/\Sigma_N \xrightarrow{d} 0$, and since $\Sigma_N/(1-\lambda)$ is asymptotically equivalent to both $\widetilde{\Sigma}_N$ and to $\sigma_{(N)}$, the theorem follows. □

Proof of Lemma 6.11:

In order to compute the kurtosis κ_N of the stationary N-day log-return $Z_N = \sum_{i=1}^N X_i$ in a GARCH(1,1) model, we first calculate the fourth moment:

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sum_{(j_1,j_2,j_3,j_4)\in\{1,\dots,N\}^4} X_{j_1}X_{j_2}X_{j_3}X_{j_4}].$$

Since for $j_1 \le j_2 \le j_3 \le j_4$ only terms with $j_1 = j_2$ and $j_3 = j_4$ can be non-zero, this can be written as

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sum_{j_1=1}^N X_{j_1}^4 + 6\sum_{j_1=1}^{N-1}\sum_{j_3=j_1+1}^N X_{j_1}^2 X_{j_3}^2]. \tag{A.15}$$

To develop this further, we need to calculate terms of the forms $\mathbb{E}[X_k^4]$ and $\mathbb{E}[X_k^2 X_{k-j}^2]$ (for $j \ge 1$) for a stationary GARCH(1,1). The first ones were computed in Lemma A.4: if $\epsilon_k \stackrel{d}{=} \epsilon$ and $\sigma_k \stackrel{d}{=} \sigma$, then $\mathbb{E}[X_k^4] = \mu_{4,\infty}$ by (A.12), which we will write as:

$$\mathbb{E}[X_k^4] = \mathbb{E}[\sigma^4]\,\mathbb{E}[\epsilon^4], \quad \mathbb{E}[\sigma^4] = \sigma_\infty^4 \frac{1-c^2}{1-c_\epsilon}. \tag{A.16}$$

Furthermore, by Lemma A.5, if $j \ge 1$,

$$\mathbb{E}[X_k^2 X_{k-j}^2] = a_0 \frac{1-c^j}{1-c}\mathbb{E}[X_{k-j}^2] + (a + \frac{b}{\mathbb{E}[\epsilon^4]})c^{j-1}\mathbb{E}[X_{k-j}^4]. \tag{A.17}$$

Combining (A.15), (A.16) and (A.17), we find that

$$\mathbb{E}[Z_N^4] = N\,\mathbb{E}[\sigma^4]\,\mathbb{E}[\epsilon^4] + 6\sum_{j_1=1}^{N-1}\sum_{j_3=j_1+1}^N a_0\frac{1-c^{j_3-j_1}}{1-c}\mathbb{E}[X_{j_1}^2] + (a+\frac{b}{\mathbb{E}[\epsilon^4]})c^{j_3-j_1-1}\mathbb{E}[X_{j_1}^4].$$

Applying (A.16) again gives

$$\mathbb{E}[Z_N^4] = N\,\mathbb{E}[\sigma^4]\,\mathbb{E}[\epsilon^4] + 6\sum_{j_1=1}^{N-1}\sum_{j_3=j_1+1}^N \sigma_\infty^2\big(1-c^{j_3-j_1}\big)\sigma_\infty^2 + (a\,\mathbb{E}[\epsilon^4]+b)c^{j_3-j_1-1}\mathbb{E}[\sigma^4],$$

which can be rewritten as

$$\begin{aligned}
\mathbb{E}[Z_N^4] &= N\,\mathbb{E}[\sigma^4]\mathbb{E}[\epsilon^4] + 6\,\sigma_\infty^4\sum_{j_1=1}^{N-1}\sum_{j_3=j_1+1}^N \big(1-c^{j_3-j_1}\big) + \\
&\quad + 6\sum_{j_1=1}^{N-1}\sum_{j_3=j_1+1}^N (a\,\mathbb{E}[\epsilon^4]+b)c^{j_3-j_1-1}\mathbb{E}[\sigma^4] \\
&= \mathbb{E}[\sigma^4]\big(N\,\mathbb{E}[\epsilon^4] + 6\,\frac{a\,\mathbb{E}[\epsilon^4]+b}{1-c}\sum_{j_1=1}^{N-1} 1-c^{N-j_1}\big) + \\
&\quad + 6\,\sigma_\infty^4\sum_{j_1=1}^{N-1}\big((N-j_1) - c\frac{1-c^{N-j_1}}{1-c}\big) \\
&= \mathbb{E}[\sigma^4]\big(N\,\mathbb{E}[\epsilon^4] + 6\,\frac{a\,\mathbb{E}[\epsilon^4]+b}{1-c}\big[(N-1) - \frac{c}{1-c}\big(1-c^{N-1}\big)\big]\Big) + \\
&\quad + 3\,\sigma_\infty^4\Big(N(N-1) - \frac{2c}{1-c}\big[(N-1) - \frac{c}{1-c}\big(1-c^{N-1}\big)\big]\Big).
\end{aligned}$$

Rearranging the terms using the expression for $\mathbb{E}[\sigma^4]$ finally yields

$$\mathbb{E}[Z_N^4] = \big(3\,N(N-1) + N\,\mathbb{E}[\epsilon^4]\frac{1-c^2}{1-c_\epsilon}\big)\,\sigma_\infty^4 + 6\,\frac{N - \frac{1-c^N}{1-c}}{1-c}\,\Big((a\,\mathbb{E}[\epsilon^4]+b)\frac{1-c^2}{1-c_\epsilon} - c\Big)\sigma_\infty^4. \tag{A.18}$$

Since, finally, $\mathbb{E}[Z_N^2] = \sum_{i=1}^N \mathbb{E}[X_i^2] = N\sigma_\infty^2$, we conclude that

$$\kappa_N = \frac{\mathbb{E}[Z_N^4]}{\mathbb{E}[Z_N^2]^2} = 3\,\frac{N-1}{N} + \mathbb{E}[\epsilon^4]\frac{1-c^2}{1-c_\epsilon}\,\frac{1}{N} + \frac{6}{N^2}\,\frac{N - \frac{1-c^N}{1-c}}{1-c}\,\Big((a\,\mathbb{E}[\epsilon^4]+b)\frac{1-c^2}{1-c_\epsilon} - c\Big). \tag{A.19}$$

□

Proof of Lemma 6.12:

We finally compute the kurtosis of $Z_N = \sum_{t=1}^N X_t$ for a stationary AR(1)-GARCH(1,1). As before,

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sum_{(j_1,j_2,j_3,j_4)\in\{1,\dots,N\}^4} X_{j_1}X_{j_2}X_{j_3}X_{j_4}].$$

Writing X_t in the form (2.4), each factor $(\sum_{j=1}^N X_j)$ can be expressed as

$$\sum_{j=1}^N X_j = \sum_{t=1}^N\sum_{k=0}^\infty \lambda^k Y_{t-k} = \sum_{i=0}^{N-1}\frac{1-\lambda^{i+1}}{1-\lambda}Y_{N-i} + \sum_{i=N}^\infty \lambda^{i-N+1}\frac{1-\lambda^N}{1-\lambda}Y_{N-i}.$$

Hence

$$\mathbb{E}[Z_N^4] = \mathbb{E}[(\sum_{i_1=0}^\infty A_{i_1}Y_{N-i_1})(\sum_{i_2=0}^\infty A_{i_2}Y_{N-i_2})(\sum_{i_3=0}^\infty A_{i_3}Y_{N-i_3})(\sum_{i_4=0}^\infty A_{i_4}Y_{N-i_4})],$$

where

$$A_i := \begin{cases} \frac{1-\lambda^{i+1}}{1-\lambda} & \text{for } i \le N-1, \\ \lambda^{i-N+1}\,\frac{1-\lambda^N}{1-\lambda} & \text{for } i \ge N. \end{cases} \tag{A.20}$$

Since for $j_1 \le j_2 \le j_3 \le j_4$ only terms with $j_1 = j_2$ and $j_3 = j_4$ can be non-zero, this can be written as

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sum_{j_1=0}^\infty (A_{j_1}Y_{N-j_1})^4 + 6\sum_{j_1=0}^\infty\sum_{j_3=j_1+1}^\infty (A_{j_1}Y_{N-j_1})^2(A_{j_3}Y_{N-j_3})^2]. \tag{A.21}$$

From (A.16) we know $\mathbb{E}[Y_k^4] = \mathbb{E}[\sigma_k^4]\mathbb{E}[\epsilon_k^4]$. For the terms $\mathbb{E}[Y_k^2Y_{k-j}^2]$ $(j \ge 1)$ we know from (A.17):

$$\mathbb{E}[Y_k^2Y_{k-j}^2] = a_0\frac{1-c^j}{1-c}\mathbb{E}[Y_{k-j}^2] + (a + \frac{b}{\mathbb{E}[\epsilon^4]})c^{j-1}\mathbb{E}[Y_{k-j}^4].$$

We now combine (A.21), (A.16) and (A.17). The first term in (A.21) can be rewritten using (A.16), the second one using (A.17):

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sigma^4]\,\mathbb{E}[\epsilon^4]\sum_{j_1=0}^{\infty} A_{j_1}^4 + \\ + 6\sum_{j_1=0}^{\infty} A_{j_1}^2 \sum_{j_3=j_1+1}^{\infty} A_{j_3}^2 \Big(a_0 \frac{1-c^{j_3-j_1}}{1-c}\mathbb{E}[Y_{N-j_1}^2] + (a + \frac{b}{\mathbb{E}[\epsilon^4]})c^{j_3-j_1-1}\mathbb{E}[Y_{N-j_1}^4]\Big).$$

Applying again (A.16) gives

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sigma^4]\,\mathbb{E}[\epsilon^4]\sum_{j_1=0}^{\infty} A_{j_1}^4 + \\ + 6\sum_{j_1=0}^{\infty} A_{j_1}^2 \sum_{j_3=j_1+1}^{\infty} A_{j_3}^2 \Big(\sigma_\infty^2(1-c^{j_3-j_1})\sigma_\infty^2 + (a\,\mathbb{E}[\epsilon^4]+b)c^{j_3-j_1-1}\mathbb{E}[\sigma^4]\Big),$$

which can be rewritten as

$$\mathbb{E}[Z_N^4] = \mathbb{E}[\sigma^4]\mathbb{E}[\epsilon^4]\sum_{j_1=0}^{\infty} A_{j_1}^4 + 6\,\sigma_\infty^4 \sum_{j_1=0}^{\infty} A_{j_1}^2 \sum_{j_3=j_1+1}^{\infty} A_{j_3}^2 + \\ + 6\left((a\,\mathbb{E}[\epsilon^4]+b)\,\mathbb{E}[\sigma^4] - c\sigma_\infty^4\right)\sum_{j_1=0}^{\infty} A_{j_1}^2 \sum_{j_3=j_1+1}^{\infty} A_{j_3}^2 c^{j_3-j_1-1},$$

with A_i as defined in (A.20). Developing this up to first order terms of λ leads to

$$\mathbb{E}[Z_N^4] = \frac{\mathbb{E}[\sigma^4]}{(1-\lambda)^4}\mathbb{E}[\epsilon^4](N-4\lambda) + 6\,\frac{\sigma_\infty^4}{(1-\lambda)^4}\Big(\frac{N(N-1)}{2} - 2\lambda(N-1)\Big) + \\ + \frac{6}{(1-\lambda)^4}\left((a\,\mathbb{E}[\epsilon^4]+b)\,\mathbb{E}[\sigma^4] - c\sigma_\infty^4\right)\cdot \\ \cdot\Big(\frac{N}{1-c} - \frac{1-c^N}{(1-c)^2} - 2\lambda\frac{1-c^{N-1}}{1-c}\Big) + O(\lambda^2).$$

Rearranging the terms using (A.16) finally yields for $a^2\mathbb{E}[\epsilon^4] + 2ab + b^2 < 1$,

$$\mathbb{E}[Z_N^4] = \big(3(N-1) + \mathbb{E}[\epsilon^4]\frac{1-c^2}{1-c_\epsilon}\big)\frac{N-4\lambda}{(1-\lambda)^4}\sigma_\infty^4 + \\ + 6\left((a\,\mathbb{E}[\epsilon^4]+b)\,\frac{1-c^2}{1-c_\epsilon} - c\right)\frac{N - \frac{1-c^N}{1-c} - 2\lambda(1-c^{N-1})}{1-c}\,\frac{\sigma_\infty^4}{(1-\lambda)^4} + O(\lambda^2). \tag{A.22}$$

Additionally, we know from (A.7) that in AR(1)-GARCH(1,1) models

$$\mathbb{E}[Z_N^2] = \frac{\sigma_\infty^2}{(1-\lambda)^2}(N - 2\lambda + O(\lambda^2)).$$

Hence,

$$\begin{aligned}\kappa_N &= \frac{\mathbb{E}[Z_N^4]}{\mathbb{E}[Z_N^2]^2} \\ &= 3\,\frac{N-1}{N} + \mathbb{E}[\epsilon^4]\frac{1-c^2}{1-c_\epsilon}\,\frac{1}{N} + \\ &\quad + \frac{6(N+4\lambda)}{N^3}\left((a\,\mathbb{E}[\epsilon^4]+b)\,\frac{1-c^2}{1-c_\epsilon} - c\right)\frac{N - \frac{1-c^N}{1-c} - 2\lambda(1-c^{N-1})}{1-c} + O(\lambda^2).\end{aligned} \tag{A.23}$$

□

Appendix B Figures

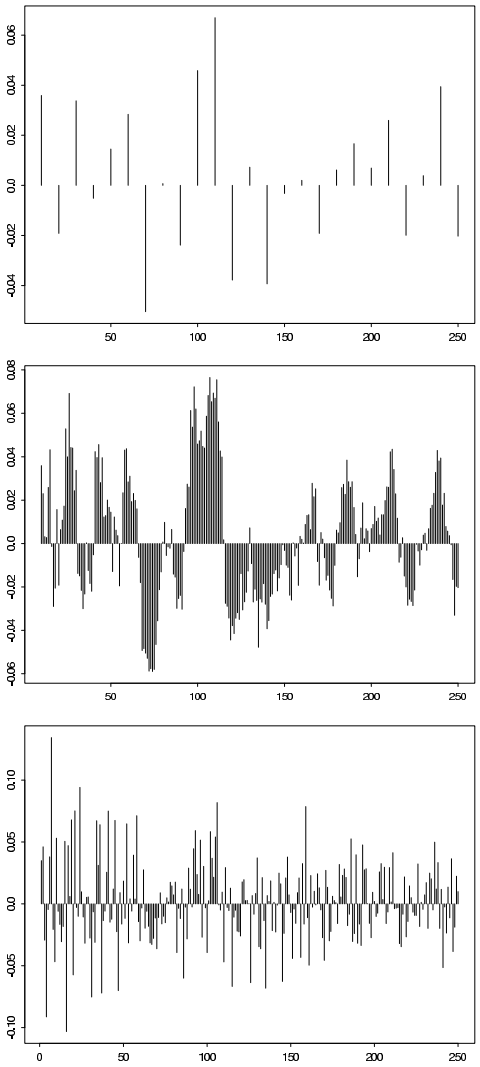

Figure 1 Top: 10-day log-returns, middle: $\sqrt{10}\cdot$1-day log-returns, bottom: overlapping 10-day log-returns. Model: GARCH(1,1).

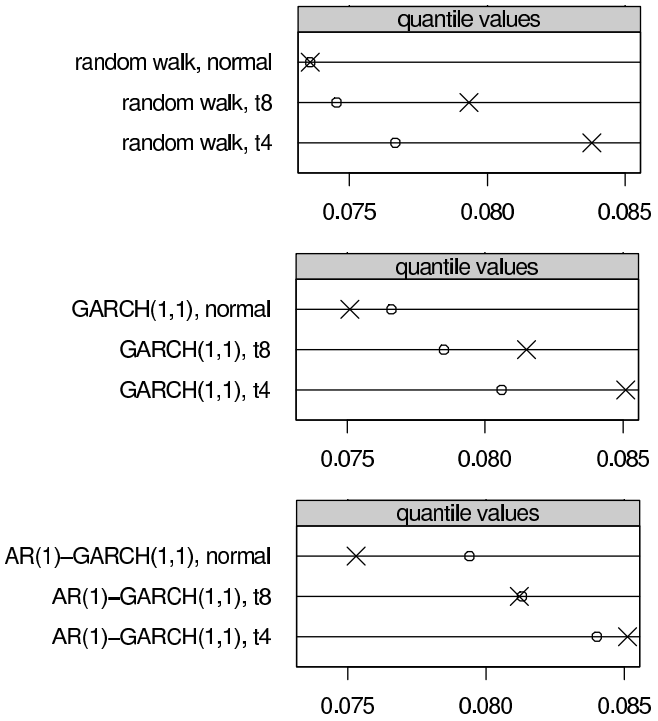

Figure 2 Estimation of $\mathrm{VaR}_{99\%}(X_t)$: circles represent the true 10-day value-at-risk, while crosses represent the scaled 1-day quantiles.

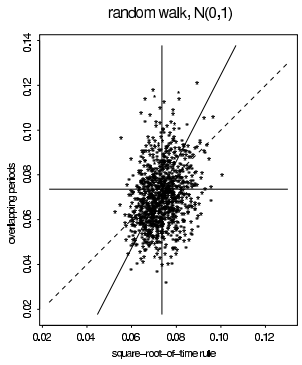

Figure 3 Comparison of two quantile estimation methods for a random walk model with normal innovations: square-root-of-time rule vs. overlapping 10-day periods.

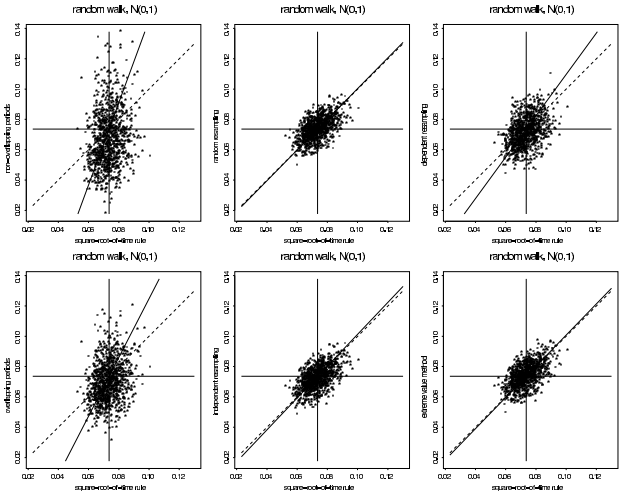

Figure 4 Comparison of the quantile estimation methods for a random walk model with normal innovations.

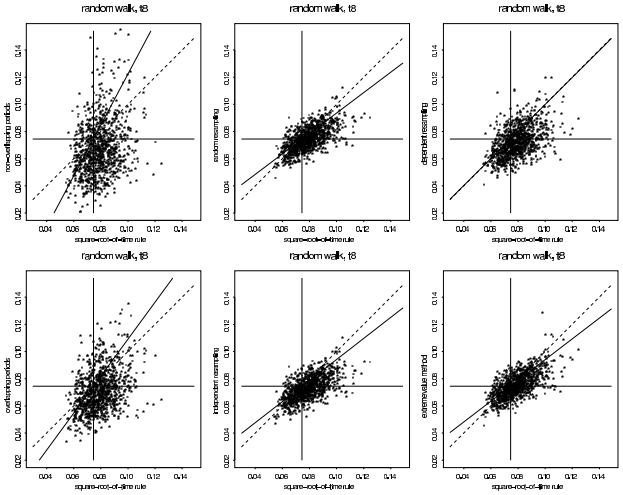

Figure 5 Comparison of the quantile estimation methods for a random walk model with Student-t_8 innovations.

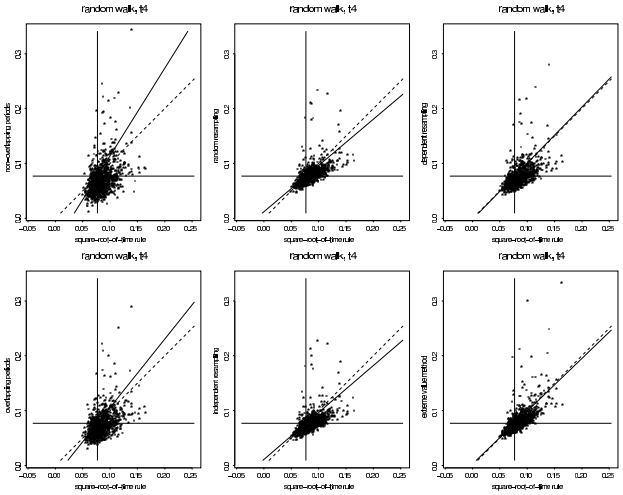

Figure 6 Comparison of the quantile estimation methods for a random walk model with Student-t_4 innovations.

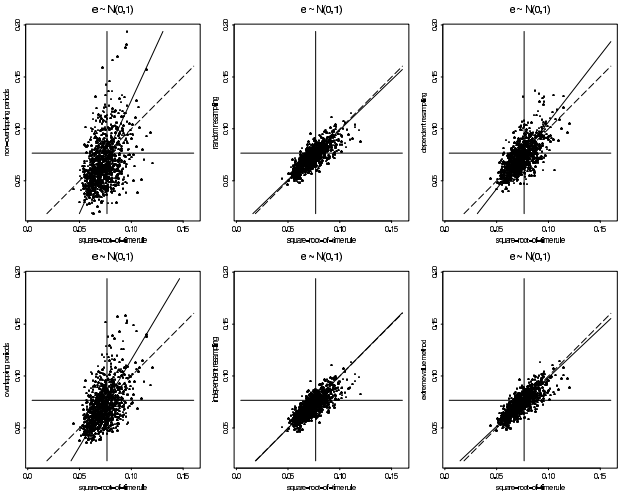

Figure 7 Comparison of the quantile estimation methods for a GARCH(1,1) model with normal innovations.

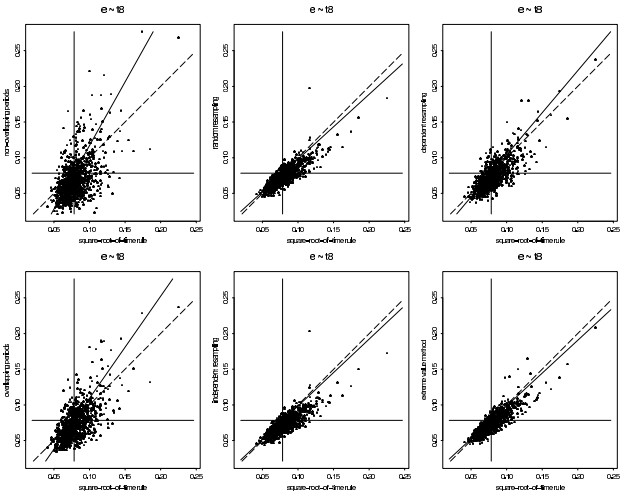

Figure 8 Comparison of the quantile estimation methods for a GARCH(1,1) model with Student-t_8 innovations.

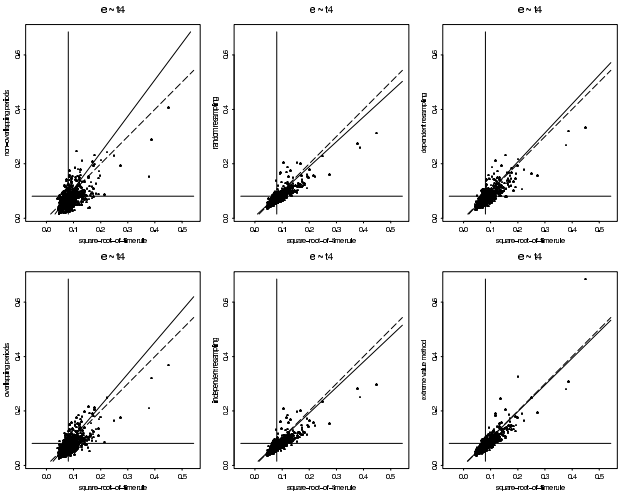

Figure 9 Comparison of the quantile estimation methods for a GARCH(1,1) model with Student-t_4 innovations.

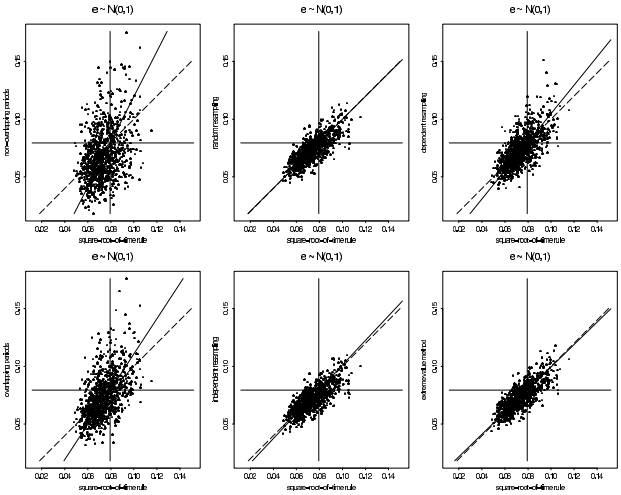

Figure 10 Comparison of the quantile estimation methods for an AR(1)-GARCH(1,1) model with normal innovations.

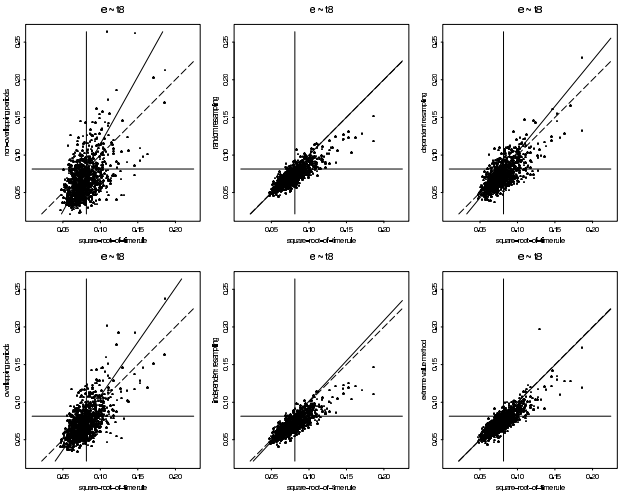

Figure 11 Comparison of the quantile estimation methods for an AR(1)-GARCH(1,1) model with Student-t_8 innovations.

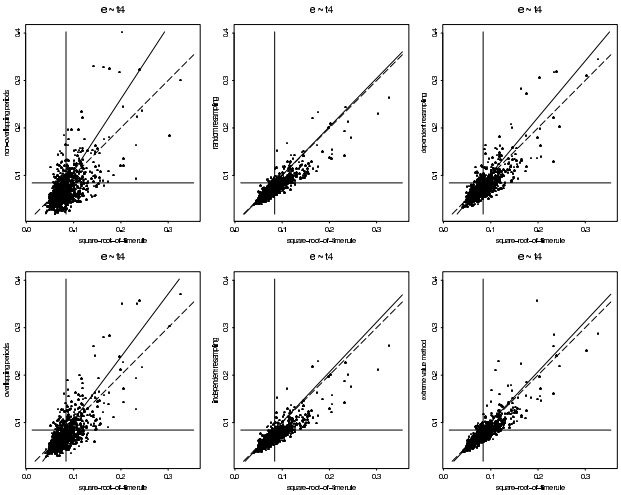

Figure 12 Comparison of the quantile estimation methods for an AR(1)-GARCH(1,1) model with Student-t_4 innovations.

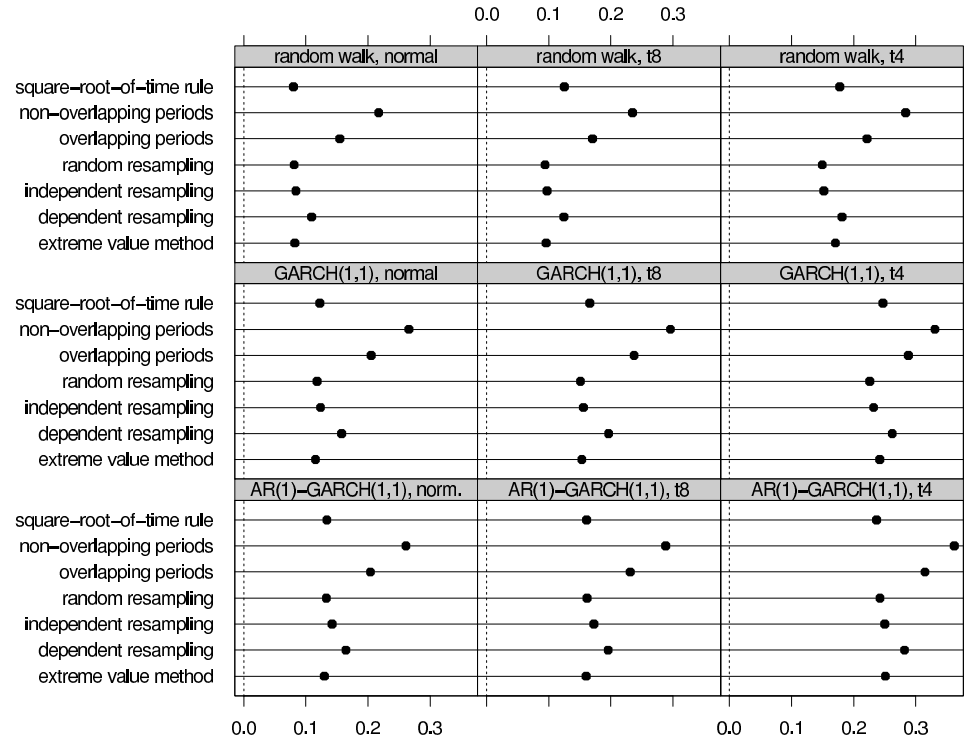

Figure 13 Mean relative deviations of the value-at-risk estimates from the true value-at-risk for the seven estimation methods in the nine models.

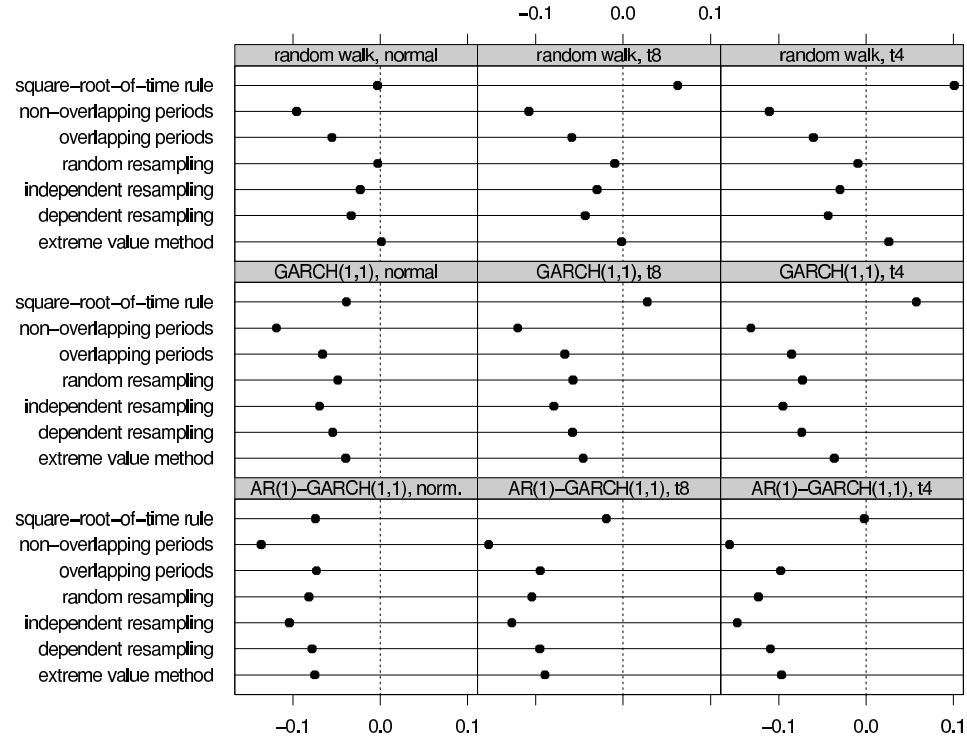

Figure 14 Relative differences between the mean of the estimated value-at-risk values, and the true value-at-risk.

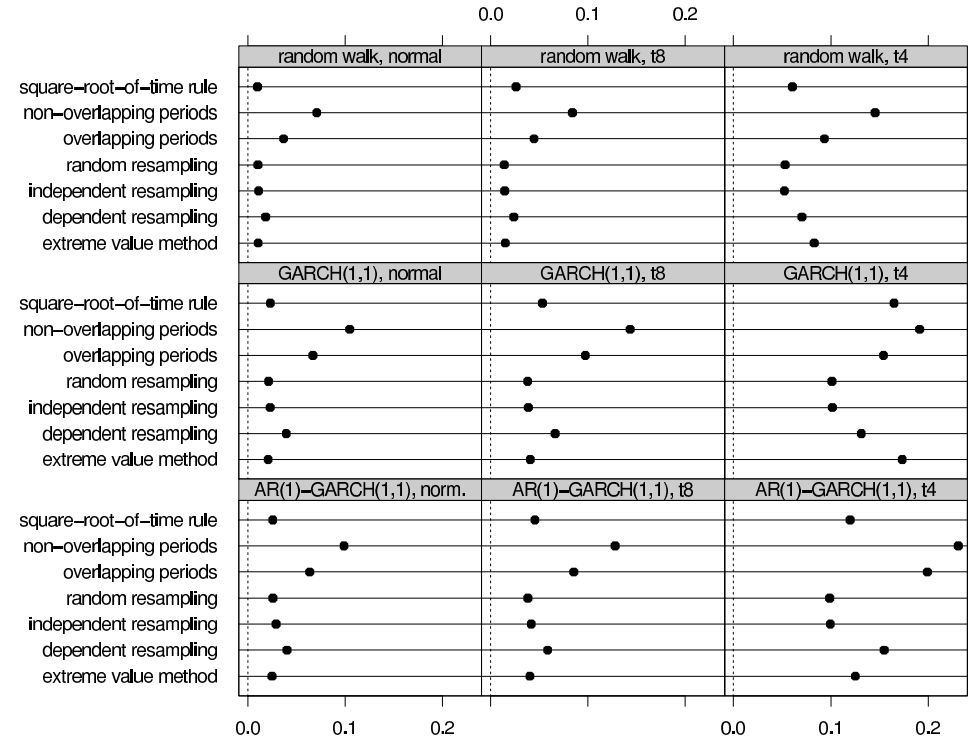

Figure 15 Mean squared error of the value-at-risk estimates.

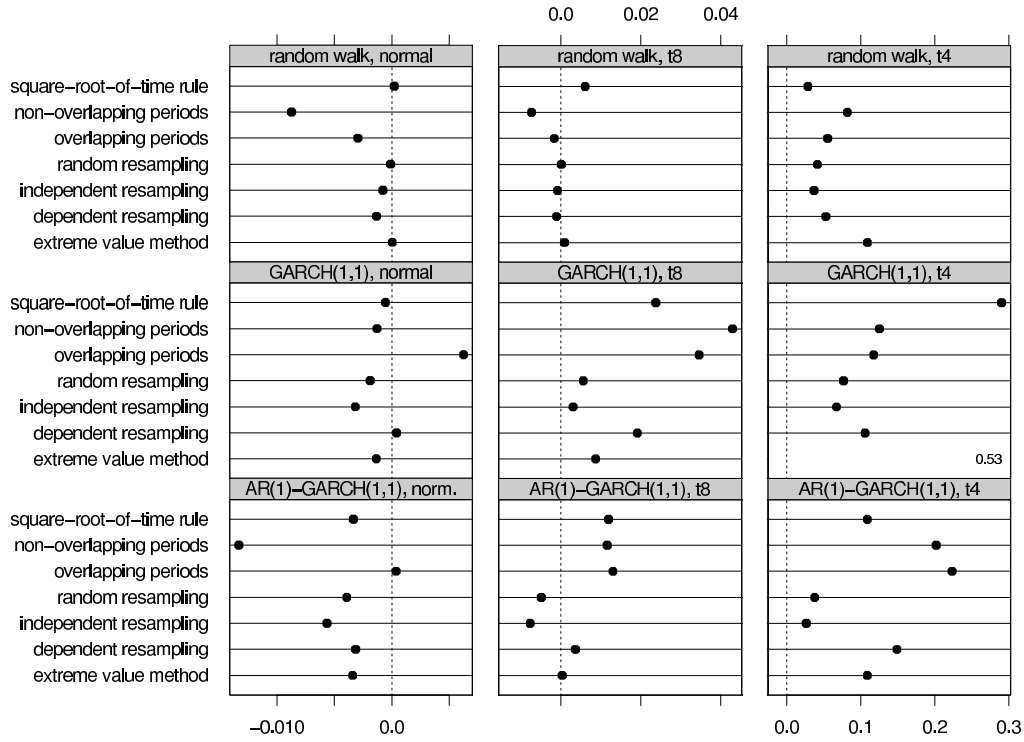

Figure 16 Evaluation of the skewness-type measure introduced in Section 3.

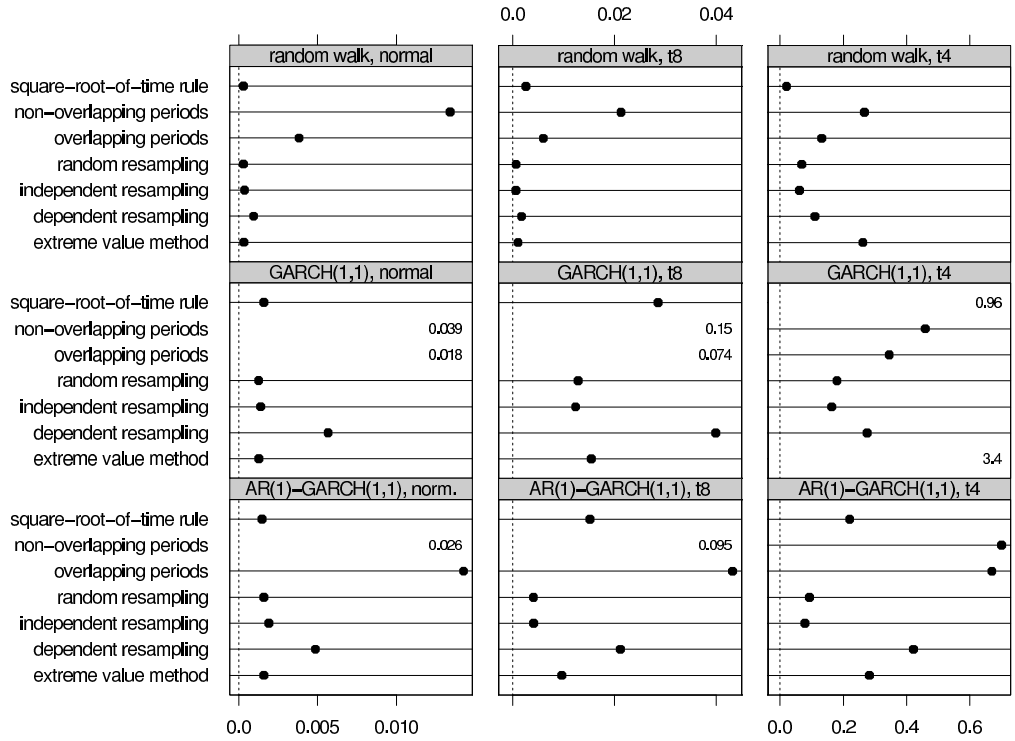

Figure 17 Evaluation of the kurtosis-type measure introduced in Section 3.

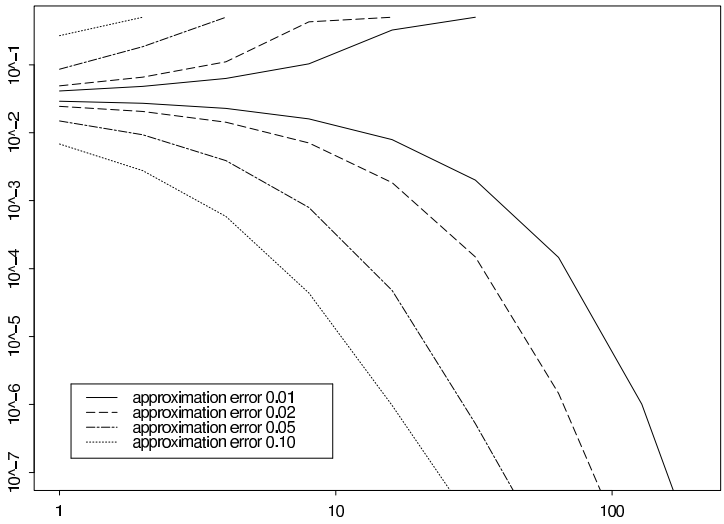

Figure 18 Contour plot for the area, where quantiles of a normal distribution are a good approximation for quantiles of the sum of Student-t_8 distributions. x-axis: number of convolutions. y-axis: one minus the level α, drawn on logarithmic scale.

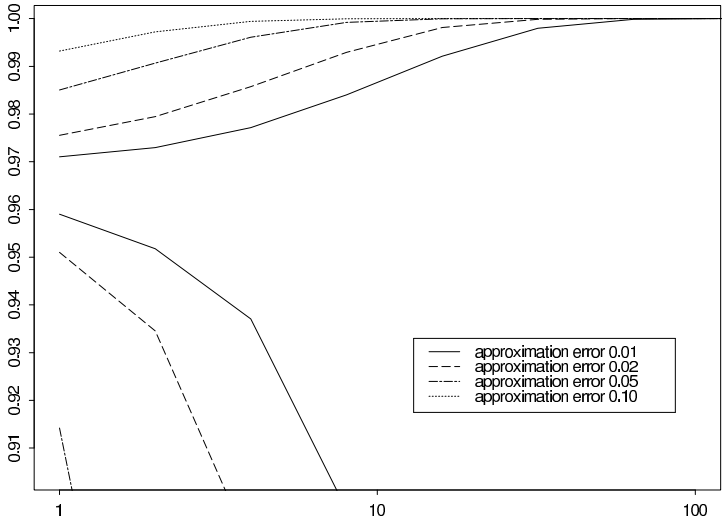

Figure 19 Contour plot for the area, where quantiles of a normal distribution are a good approximation for quantiles of the sum of Student-t_8 distributions. x-axis: number of convolutions. y-axis: level α, drawn on linear scale.

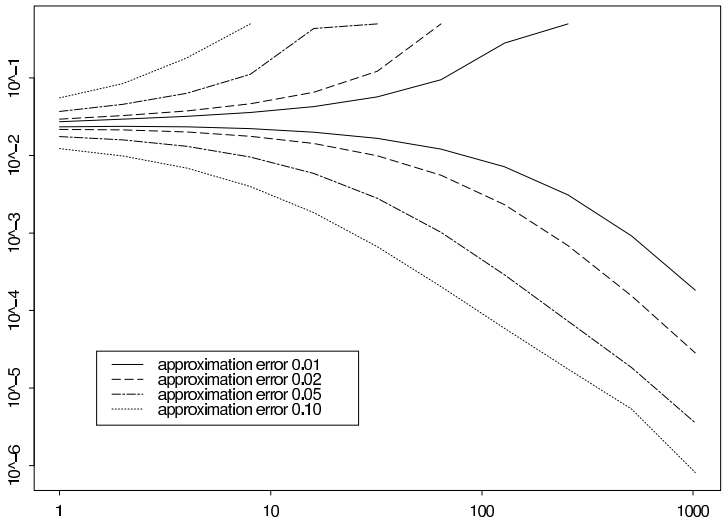

Figure 20 Contour plot for the area, where quantiles of a normal distribution are a good approximation for quantiles of the sum of Student-t_4 distributions. x-axis: number of convolutions. y-axis: one minus the level α, drawn on logarithmic scale.

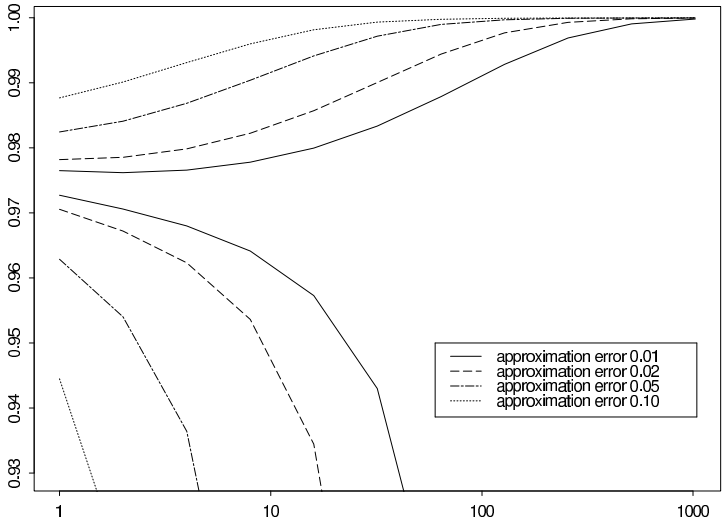

Figure 21 Contour plot for the area, where quantiles of a normal distribution are a good approximation for quantiles of the sum of Student-t_4 distributions. x-axis: number of convolutions. y-axis: level α, drawn on linear scale.

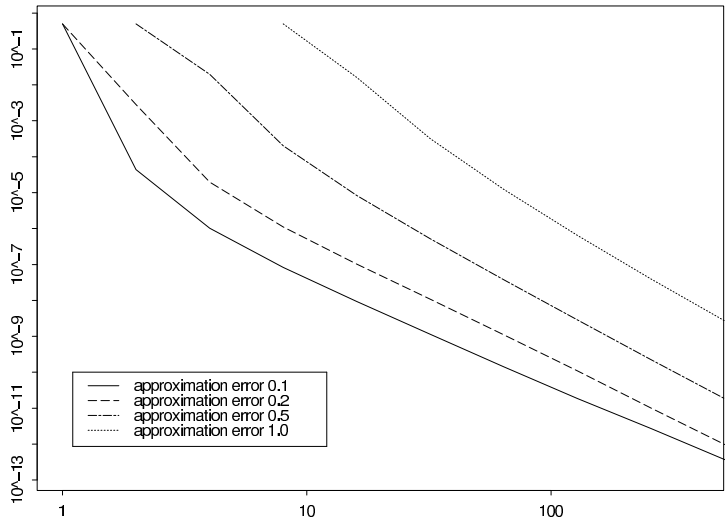

Figure 22 Contour plot with the thresholds, from where on quantiles of a Student-t_8 distribution are a good approximation for quantiles of the sum of Student-t_8 distributions. x-axis: number of convolutions. y-axis: one minus the level α, drawn on logarithmic scale.

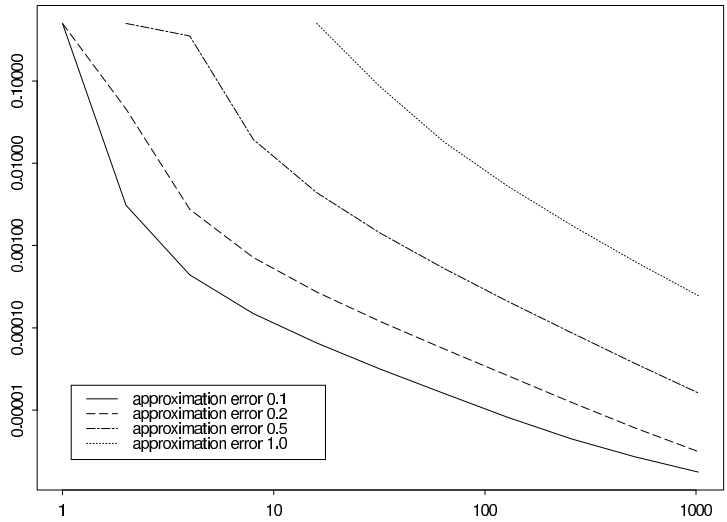

Figure 23 Contour plot with the thresholds, from where on quantiles of a Student-t_4 distribution are a good approximation for quantiles of the sum of Student-t_4 distributions. x-axis: number of convolutions. y-axis: one minus the level α, drawn on logarithmic scale.

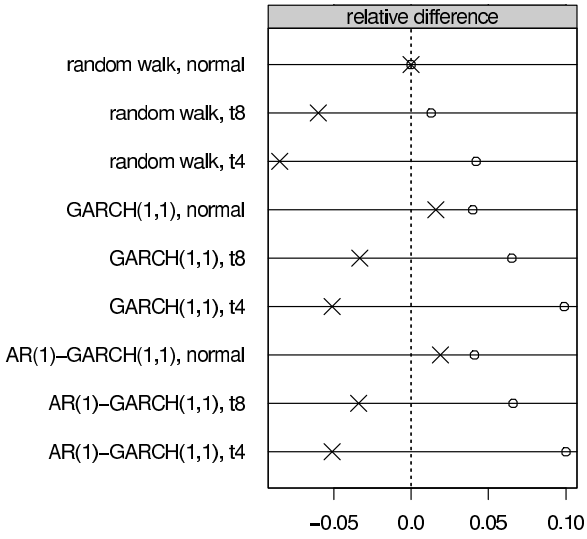

Figure 24 For each model we evaluate the relative differences between the standardised 99% 10-day quantiles and two different reference quantiles: on the one hand the reference quantile is the 99% quantile of a standard normal distribution (circles), on the other hand we take the standardised 99% quantile of the one-day model as reference quantile (crosses).

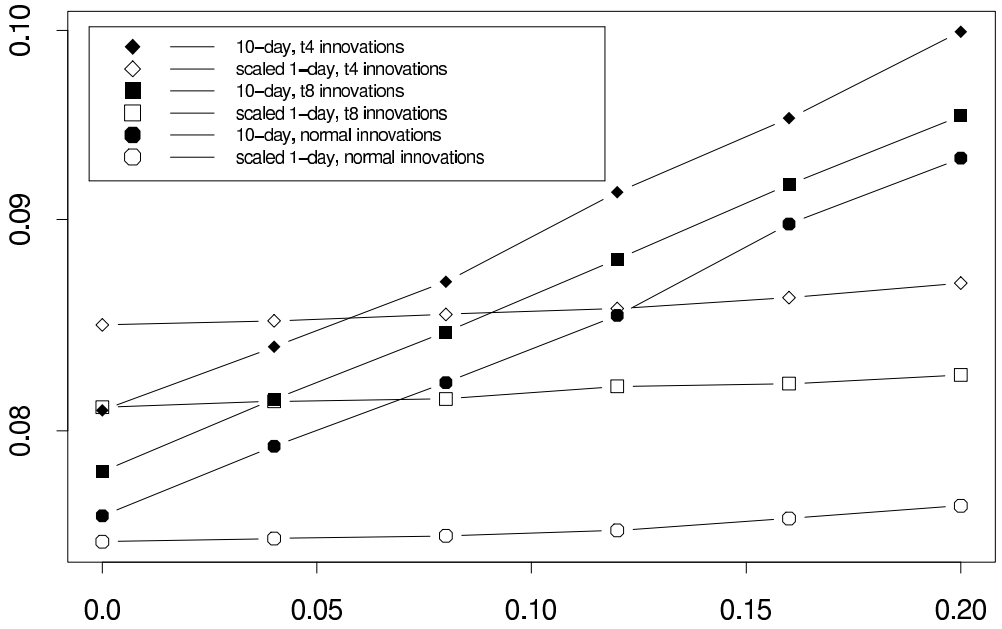

Figure 25 Simulated values for the 10-day 99% value-at-risk in AR(1)-GARCH(1,1) models: true VaR (black symbols) and VaR using the $\sqrt{10}$-rule (white symbols). Parameters: $a = 0.05$, $b = 0.92$, $\lambda \in [0.00, 0.20]$.

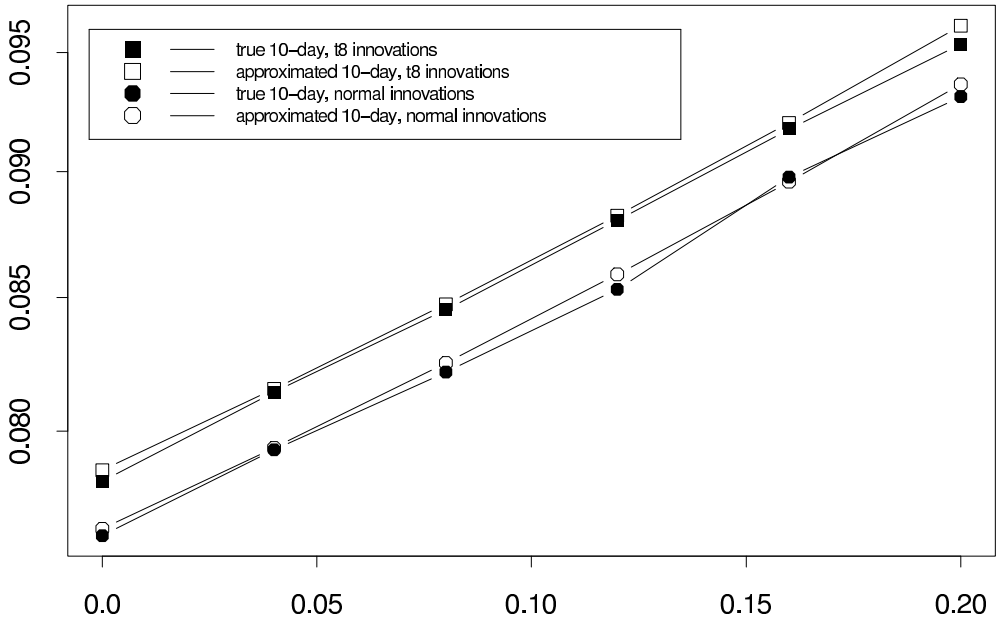

Figure 26 Simulated (black symbols) versus approximated (white symbols) values for the 10-day 99% value-at-risk in AR(1)-GARCH(1,1) models. Parameters: $a = 0.05$, $b = 0.92$, $\lambda \in [0.00, 0.20]$.

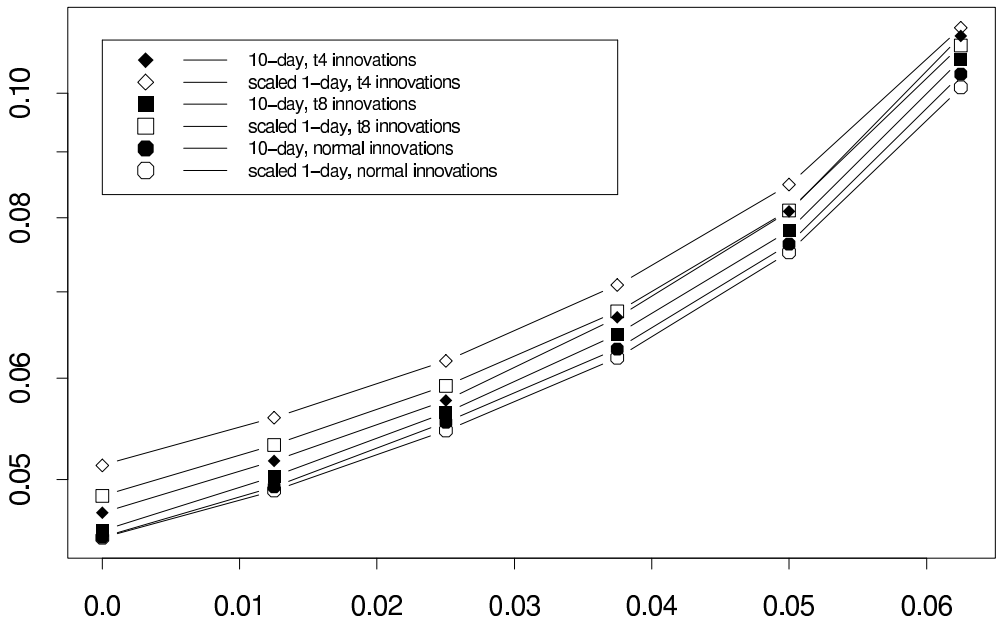

Figure 27 Simulated values for the 10-day 99% value-at-risk in GARCH(1,1) models: true VaR (black symbols) and VaR using the $\sqrt{10}$-rule (white symbols). Parameters: $b = 0.92$, $a \in [0.00, 0.0625]$.

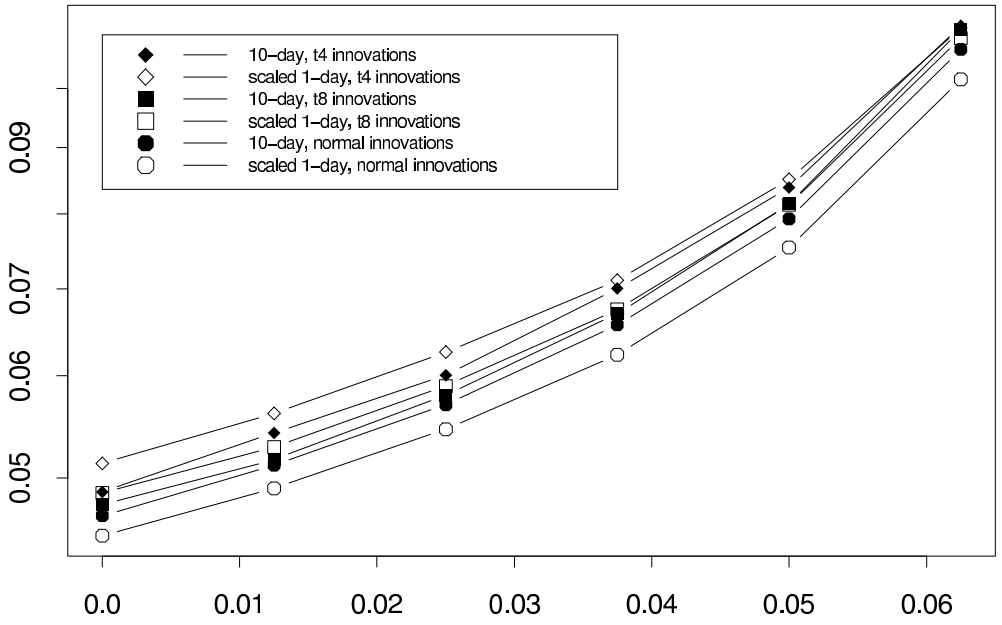

Figure 28 Simulated values for the 10-day 99% value-at-risk in AR(1)-GARCH(1,1) models: true VaR (black symbols) and VaR using the $\sqrt{10}$-rule (white symbols). Parameters: $\lambda = 0.04$, $b = 0.92$, $a \in [0.00, 0.0625]$.

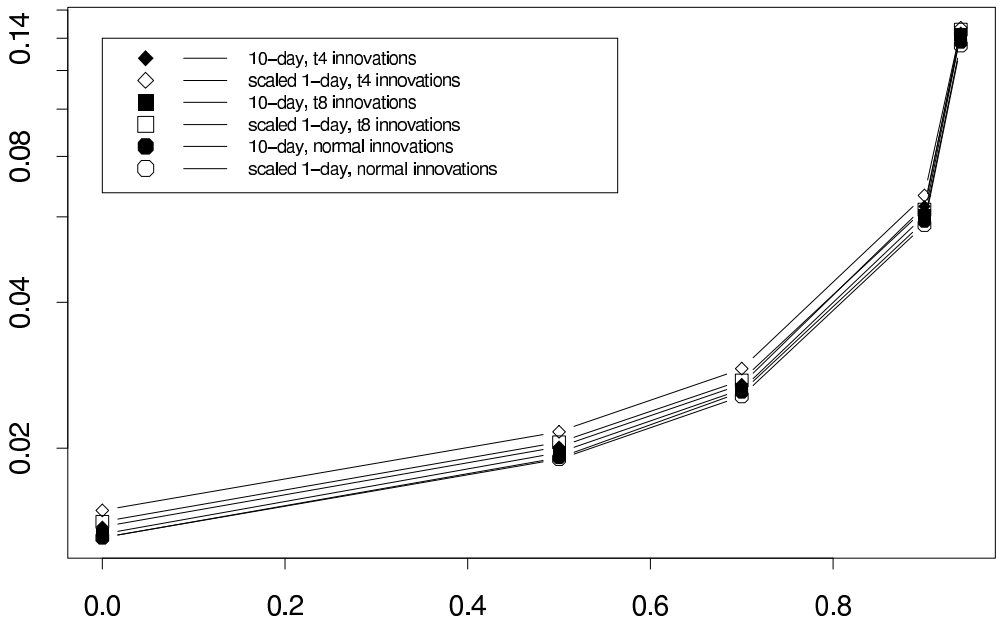

Figure 29 Simulated values for the 10-day 99% value-at-risk in GARCH(1,1) models: true VaR (black symbols) and VaR using the $\sqrt{10}$-rule (white symbols). Parameters: $a = 0.05$, $b \in [0.00, 0.94]$.

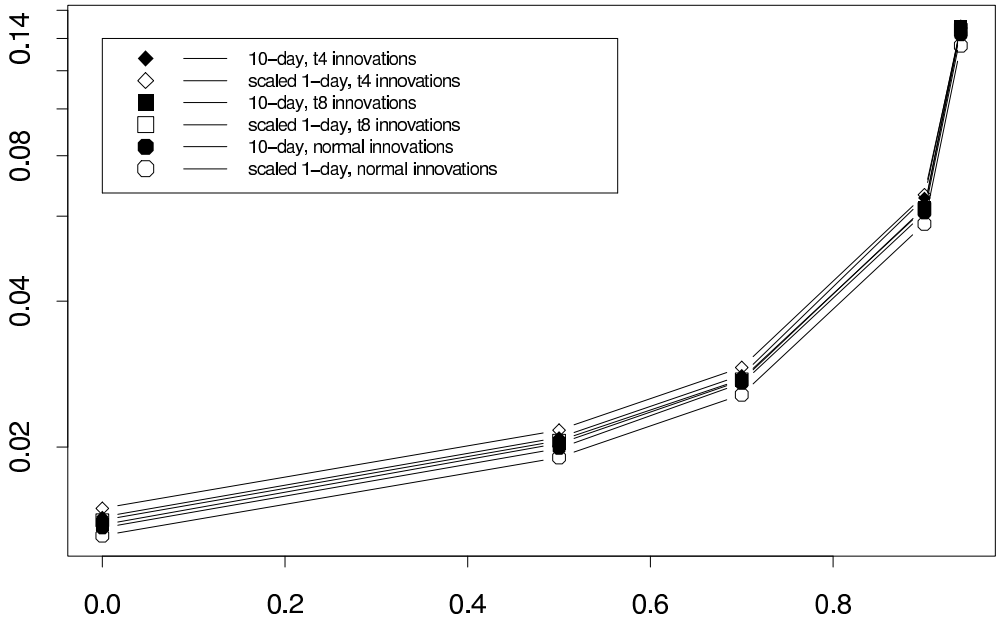

Figure 30 Simulated values for the 10-day 99% value-at-risk in AR(1)-GARCH(1,1) models: true VaR (black symbols) and VaR using the $\sqrt{10}$-rule (white symbols). Parameters: $\lambda = 0.04$, $a = 0.05$, $b \in [0.00, 0.94]$.