



---

No. 2006/23

**Accurate Value-at-Risk Forecasting Based on  
the (good old) Normal-GARCH Model**

Christoph Hartz, Stefan Mittnik,  
and Marc S. Paoletta

---





## Center for Financial Studies

The *Center for Financial Studies* is a nonprofit research organization, supported by an association of more than 120 banks, insurance companies, industrial corporations and public institutions. Established in 1968 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia.

The CFS Working Paper Series presents the result of scientific research on selected topics in the field of money, banking and finance. The authors were either participants in the Center's Research Fellow Program or members of one of the Center's Research Projects.

If you would like to know more about the *Center for Financial Studies*, please let us know of your interest.

Prof. Dr. Jan Pieter Krahen

Prof. Volker Wieland, Ph.D.



CFS Working Paper No. 2006/23

**Accurate Value-at-Risk Forecast with the  
(good old) Normal-GARCH Model\***

Christoph Hartz<sup>1</sup>, Stefan Mittnik<sup>2</sup>,  
and Marc S. Paolella<sup>3</sup>

October 2006

**Abstract:**

A resampling method based on the bootstrap and a bias-correction step is developed for improving the Value-at-Risk (VaR) forecasting ability of the normal-GARCH model. Compared to the use of more sophisticated GARCH models, the new method is fast, easy to implement, numerically reliable, and, except for having to choose a window length  $L$  for the bias-correction step, fully data driven. The results for several different financial asset returns over a long out-of-sample forecasting period, as well as use of simulated data, strongly support use of the new method, and the performance is not sensitive to the choice of  $L$ .

**JEL Classification:** C22, C53, C63, G12

**Keywords:** Bootstrap, GARCH, Value-at-Risk

\* The authors wish to thank Helmut Herwartz for helpful discussions on an earlier draft of this paper, as well as the excellent comments and suggestions from two anonymous referees.

1 Department of Statistics, University of Munich

2 Department of Statistics, University of Munich, Center for Financial Studies, Frankfurt, and Ifo Institute for Economic Research, Munich

3 Corresponding author: Swiss Banking Institute, University of Zurich, Switzerland, E-mail: paolella@isb.unizh.ch

## Non-technical summary

Despite its widespread use in empirical financial analysis the normal distribution is, in general, not suitable for modeling returns on financial assets. In particular, the well known phenomenon of time-varying volatility of financial returns requires a conditional approach to modeling asset returns. The most prominent model for capturing the time varying volatility is the generalized autoregressive conditional heteroskedasticity (GARCH) model. But even if time-varying volatility is taken into account, i.e., asset returns are conditionally normal and heteroskedasticity is captured by a GARCH model, GARCH-filtered returns are not adequately described by a normal distribution. Especially the tails of the return distribution are poorly approximated by the normal distribution, resulting in biased Value-at-Risk (VaR) forecasts of the normal-GARCH model. Ways for producing better VaR forecasts is to use more sophisticated GARCH dynamics and/or allow for other distributions. The downside of this strategy is twofold. Stationarity conditions and other properties of interest of more sophisticated models can be very complex compared to those of the plain normal-GARCH model; and the practical implementation of some of the proposed models is (most-often) nontrivial.

This paper presents a resampling method based on the bootstrap and a bias-correction step is developed for improving the VaR forecasting ability of the normal-GARCH model. Compared to the use of more sophisticated GARCH models, the new method is fast, easy to implement, numerically reliable, and, except for having to choose a window length  $L$  for the bias-correction step, fully data driven. The results for several different financial asset returns over a long out-of-sample forecasting period as well as use of simulated data strongly support the new method.

## Nichttechnische Zusammenfassung

Die Normalverteilung ist, entgegen ihrer hohen Verbreitung in der empirischen Finanzanalyse, im allgemeinen nicht dazu geeignet, die Renditen von Finanzmarkt-Zeitreihen adäquat zu beschreiben. Ein viel beobachtetes Phänomen ist insbesondere die über die Zeit variierende Volatilität der Renditen, die eine bedingte Modellierung der Renditen notwendig erscheinen läßt. Der wohl am weitesten verbreitete Ansatz um solche Volatilitätsschwankungen zu modellieren ist das GARCH-Modell. Doch auch bei Berücksichtigung der Volatilitätsschwankungen, d.h. bei bedingter Modellierung der Renditen mit Hilfe eines GARCH-Modells, ist die Normalverteilung im allgemeinen nicht dazu geeignet, die Verteilung der GARCH-gefilterten Renditen ausreichend genau zu beschreiben. Insbesondere Value-at-Risk (VaR) Prognosen sind mit dem normal-GARCH Modell im allgemeinen verzerrt, da die Normalverteilung die Enden der Rendite-Verteilung nur unzureichend beschreibt. Mögliche Auswege scheinen die Erweiterung und Modifikation der GARCH Dynamik, sowie die Verwendung anderer Verteilungen. Dies führt jedoch im allgemeinen dazu, daß diese Modelle sowohl theoretisch, als auch praktisch schwerer zu beherrschen sind.

In der vorliegenden Studie entwickeln wir eine auf dem Bootstrap basierende Methode mit einem Verzerrungs-Korrektur Schritt, um die VaR Prognoseeigenschaften des normal-GARCH Modells zu verbessern. Im Vergleich zur Verwendung von komplexeren GARCH Spezifikationen und/oder Verteilungsannahmen ist diese neue Methode schnell, einfach zu implementieren, numerisch zuverlässig und (abgesehen von einer zu wählenden Fensterlänge  $L$  für den Schritt zur Korrektur der VaR-Verzerrung) vollständig Daten getrieben.

Die vorgeschlagene Methode wird in langen out-of-sample Prognosezeiträumen auf ihre VaR Prognosefähigkeiten geprüft. Sowohl für verschiedene Finanzmarkt-Reihen, als auch für simulierte Daten, erweist sich die neue Methode als sehr gut geeignet, die VaR Prognosen der normal-GARCH Modells entscheidend zu verbessern und liefert auch im Vergleich zu komplexeren Modellen sehr gute Ergebnisse.

# 1 Introduction

The Value-at-Risk, or VaR, has established itself as the most prominent measure of financial downside market risk. While it has been criticized as being theoretically deficient (e.g., non-subadditive) and numerically problematic (it is nonconvex) (see Dowd and Blake, 2006, for an introduction, survey, and original references), it is also still recognized—despite its drawbacks compared to coherent risk measures—as the most widely used risk measure in practice, and its accurate computation is also fundamental for the computation of other quantile-based risk measures such as expected shortfall (Dowd and Blake, 2006, page 194).

Owing to the Basle Committee’s (1995, 1996) Internal Model Approach, which allows banks to implement in-house VaR models for calculating capital requirements, the number of methods for such calculations continues to increase. The theoretical and computational complexity of these approaches, however, is also increasing. Examples include the use of extreme value theory (McNeil and Frey, 2000), quantile regression methods (Engle and Manganelli, 2004), and Markov switching techniques (Gray, 1996; Klaassen, 2002; Haas, Mittnik and Paoletta, 2004b); see Kuuster, Mittnik and Paoletta (2005) and the references therein for an overview and comparisons of these and further models.

While greatly differing in approach, all these methods are able to account for the two most conspicuous characteristics of financial asset returns, namely strong time-varying volatility and excess kurtosis relative to the normal distribution. Simplistic methods, such as using the empirical distribution function of a moving window of returns to compute the tail quantiles (often referred to, somewhat misleadingly, as historical simulation), cannot adequately account for the volatility clustering and perform very poorly in practice. Nevertheless, it is computationally trivial, with simple and well-known statistical properties (under an iid assumption), whereas the aforementioned methods are far from trivial, both in terms of numerical procedures for estimation and statistical inference.

The least sophisticated method which can still capture the two primary stylized facts to a reasonable extent is the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986). In general, for the set of equally-spaced asset returns,  $r_t$ ,  $t = 1, \dots, T$ , the class of ARMA( $p, q$ )–GARCH( $r, s$ ) models is given by

$$r_t = a_0 + \sum_{i=1}^p a_i r_{t-i} + \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}, \quad \epsilon_t = z_t \sigma_t, \quad (1)$$

$$\sigma_t^2 = c_0 + \sum_{i=1}^r c_i \epsilon_{t-i}^2 + \sum_{j=1}^s d_j \sigma_{t-j}^2, \quad (2)$$

where the  $z_t$ ,  $t = 1, \dots, T$ , are independent and identically distributed (iid) standard normal random variables. Particularly for the most common formulation with  $p = 1$ ,  $q = 0$ ,  $r = s = 1$ , it has the advantage of being (relative to the aforementioned methods) very simple to estimate, and its theoretical properties of interest, such as moments and stationarity conditions, are tractable.

The inadequacy of the normal–AR(1)–GARCH(1,1) model for in-sample fit and out-of-sample forecasting became obvious not long after its inception, and was superseded by replacing

the normal assumption by the Student's  $t$ , whereby the degrees of freedom parameter is interpreted as an additional distributional shape parameter in  $\mathbb{R}_+$  and is estimated jointly with the location and scale model parameters given in the recursions (1) and (2). While better than a normal-GARCH model, particularly for more extreme (1% or less) VaR thresholds, the  $t$ -GARCH can also be improved upon by generalizing both the parametric form of the time-varying volatility in (2) and the distributional assumption (Mittnik and Paoletta, 2000, 2003; Giot and Laurent, 2004).

There now exist a wide variety of generalizations of the functional form in (2), and a large number of candidate distributions for the innovation sequence, several combinations of which have been shown to be very capable of capturing the various empirical features of the returns and also for delivering accurate out-of-sample predictions of the entire distribution of a future return or just particular quantiles, as is needed for VaR-prediction (see Alexander, 2001, Ch. 9 and 10; Bao, Lee and Saltoglu, 2003; and the references therein). The downside of this development is twofold. Firstly, stationarity conditions (and other properties of interest such as the moments) of more sophisticated models are very complex compared to those of the plain normal-GARCH model (see, e.g., He, Teräsvirta, and Malmsten, 2002; Mittnik, Paoletta and Rachev, 2002; Karanasos and Kim, 2003; and Haas, Mittnik and Paoletta 2004a,b).

Secondly, the numeric implementation of some of the proposed models is nontrivial. For example, the quadratic GARCH model of Sentana (1995) has been demonstrated by several authors to provide better forecasts than competing models and better “whitening” or signal extraction from the data (see Paoletta, 2001; Mittnik, Paoletta, and Rachev, 2000; and the references therein), but has the drawbacks of requiring a relatively large number of parameters and can result in negative scale parameters, both of which exacerbate the numeric computation of the maximum likelihood estimate, and bars use of less sophisticated software such as spreadsheets and “menu driven” econometrics packages. Similarly, the EGARCH model introduced by Nelson (1991), which possesses some theoretical advantages over GARCH model (2), is known to be very problematic in practice, with the choice of starting values being extremely critical for successful likelihood maximization (Frachot, 1995; Franses and van Dijk, 1996).

A similar critique applies to the distributional assumption, in that the density (required for the likelihood function) and distribution function (for computing the VaR) may not be expressible in closed form. Examples include the hyperbolic distribution and Gauss-Laplace mixtures (Haas, Mittnik and Paoletta, 2006), noncentral Student's  $t$  (Harvey and Siddique, 1999; Broda and Paoletta, 2006), geometric stable and stable Paretian (Rachev and Mittnik, 2000). These require nontrivial numeric procedures such as numeric integration, special function libraries, fast Fourier transform methods, multivariate root-finding, etc., which are not germane to many software platforms, and possibly not familiar to applied practitioners.

The goal of this paper is to recover the simplicity of the normal-AR(1)-GARCH(1,1) model, but still overcome its deficiencies with respect to VaR forecasting. To this end, we propose a data driven method based on it, and use resampling methods to correct for the clear tendency of the model to underestimate the VaR. While resampling methods are, in general, “numerically

intensive”, it is, in this case, very simple to implement, and is actually faster to estimate than several, more complicated models. It is also numerically extremely reliable, thus obviating the need for advanced numerical methods, user intervention, carefully selected starting values, etc., and also bypasses the inherent theoretical complexity of the more advanced models. Our results are quite encouraging and demonstrate that the simple normal–GARCH model need not be abandoned after all.

The remainder of this paper is as follows. Section 2 discusses the methodology for approximating the distribution of the VaR point forecast and how it can be used to improve its accuracy, while Section 3 illustrates the method using several financial time series as well as simulated data and demonstrates its effectiveness in improving the accuracy of out-of-sample VaR predictions. Section 4 provides some concluding remarks.

## 2 Forecasting Value-at-Risk

Interest centers on using the past returns,  $r_t$ ,  $t = 1, \dots, T$ , formed from an equally-spaced time series of prices of a financial asset, to estimate a particular quantile,  $\kappa_\lambda = \kappa_\lambda(h, T)$ , of the predictive distribution of  $r_{T+h|T}$  for a time horizon  $h \in \mathbb{Z}$  (usually 1 day or 1 week), and a given probability level  $\lambda$  (with usual values between and including 1% to 5%).

For a given return series and a chosen model from the normal–ARMA–GARCH class in (1) and (2), the usual conditional VaR forecast is obtained by estimating the unknown parameter vector  $\boldsymbol{\theta} = (a_0, \dots, a_p, b_1, \dots, b_q, c_0, \dots, c_r, d_1, \dots, d_s)'$  via conditional maximum likelihood. Note: The usual conditioning is used, i.e.,  $r_1, \dots, r_p$  are assumed to be fixed values so that the effective sample size is  $T - p$  and  $\epsilon_0, \dots, \epsilon_{1-q}$  are set to zero. The start-up values  $\sigma_0^2, \dots, \sigma_{1-s}^2$  can be set either to their unconditional (estimated) values of  $\hat{c}_0 / (1 - \sum_i \hat{c}_i - \sum_j \hat{d}_j)$  or to the sample variance of the returns, the latter being preferred because of the tendency for  $\sum \hat{c}_i + \sum \hat{d}_j$  to be close to or exceed one. Values  $\epsilon_0^2, \dots, \epsilon_{1-r}^2$  are set to  $E[z_t^2] \sigma_0^2 = \sigma_0^2$ .

This estimator is consistent under standard regularity conditions (see, e.g., Gouriéroux, 1997), so even for non-normal innovations with existing second moment, we obtain a quasi maximum likelihood (QML) parameter vector estimate  $\hat{\boldsymbol{\theta}}$ .

In addition to the point estimator  $\hat{\boldsymbol{\theta}}$ , we also define the set of estimated standardized residuals  $\{\hat{z}_t\}$ ,  $t = p, \dots, T$ , as  $\hat{z}_t = \hat{\epsilon}_t / \hat{\sigma}_t$ , with

$$\hat{\epsilon}_t = r_t - \hat{a}_0 - \left( \sum_{i=1}^p \hat{a}_i r_{t-i} \right) - \left( \sum_{j=1}^q \hat{b}_j \hat{\epsilon}_{t-j} \right), \quad \hat{\sigma}_t^2 = \hat{c}_0 + \sum_{i=1}^r \hat{c}_i \hat{\epsilon}_{t-i}^2 + \sum_{j=1}^s \hat{d}_j \hat{\sigma}_{t-j}^2, \quad (3)$$

and the  $h$ -step-ahead VaR forecast

$$\hat{\kappa}_\lambda = \hat{\kappa}_\lambda(h, T) = \Phi^{-1}(\lambda; \hat{\mu}_{T+h}, \hat{\sigma}_{T+h}^2), \quad (4)$$

where  $\Phi^{-1}(\lambda; \mu, \sigma^2)$  denotes the inverse cdf of the standard normal distribution with mean  $\mu$  and variance  $\sigma^2$ , and

$$\hat{\mu}_{T+h} = \hat{a}_0 + \sum_{i=1}^p \hat{a}_i r_{T+h-i} + \sum_{j=1}^q \hat{b}_j \hat{\epsilon}_{T+h-j}, \quad \hat{\sigma}_{T+h}^2 = \hat{c}_0 + \sum_{i=1}^r \hat{c}_i \hat{\epsilon}_{T+h-i}^2 + \sum_{j=1}^s \hat{d}_j \hat{\sigma}_{T+h-j}^2, \quad (5)$$



obtained in a natural way by iterating (1) and (2) and replacing unobserved values of  $r_\tau$ ,  $\hat{\epsilon}_\tau$  and  $\hat{\epsilon}_\tau^2$  by their conditional expectation, i.e., for  $\tau > T$ ,  $r_\tau = \mathbb{E}[r_\tau|T] = \hat{\mu}_\tau$ ,  $\hat{\epsilon}_\tau = \mathbb{E}[\hat{\epsilon}_\tau|T] = 0$  and  $\hat{\epsilon}_\tau^2 = \mathbb{E}[\hat{\epsilon}_\tau^2|T] = \hat{\sigma}_\tau^2$ .

## 2.1 Value-at-Risk Forecast Distribution

In addition to the VaR point estimator  $\hat{\kappa}_\lambda$ , our bias-correction method requires (an approximation to) the corresponding sampling distribution, for which a standard bootstrap method suggests itself. Our method coincides with that described in Pascual, Romo and Ruiz (2006) and also proposed and used by Christoffersen and Goncalves (2005) for constructing confidence intervals of VaR forecasts (see also Giamouridis, 2006). It is also related to the filtered historical simulation method studied in Barone-Adesi, Giannopoulos and Vosper (1999, 2002) and the bootstrap methodology discussed in Dowd (2005, Chapter 4).

The intuition behind the use of the bootstrap is straightforward and explains why its use has been independently proposed by several authors: The sampling distribution of the point forecast  $\hat{\kappa}_\lambda(h, T)$  is unknown and analytically intractable, but the bootstrap resampling algorithm allows a computationally simple and feasible method of approximating it. This is outlined below. Under the usual assumption that the true data generating process is constant over (at least a reasonably-sized window of) time, the bias induced by the use of the wrong, but simple, normal-GARCH model will exhibit certain regularities and thus can be corrected based on a set of past VaR bootstrap distributions. This is the assumption used in our bias-correction method outlined in Section 2.2 below, and, anticipating our positive empirical results, appears tenable in this context.

The resampling method requires drawing from the filtered (and, thus, approximately iid) innovations  $\{\hat{z}_t\}$ . Using  $B$  bootstrap replications, the  $b$ th replication,  $b = 1, \dots, B$ , entails the following steps, which are also illustrated in Figure 1 as a process chart.

**Step 0:** For a chosen set of values  $p, q, r, s$  (for which  $p = r = s = 1$  and  $q = 0$  is most common), obtain QML parameter vector estimate  $\hat{\theta}$ , estimated standardized residuals  $\{\hat{z}_t\}$ , and, for a given  $h$  (we use  $h = 1$  throughout), VaR forecast  $\hat{\kappa}_\lambda(h, T)$  from (4).

**Step 1:** Simulate the  $(b')$ th of  $B$  normal-ARMA( $p, q$ )-GARCH( $r, s$ ) time series,  $\{r_t^{(b')}\}$ , of length  $T$ , using the estimated parameter vector  $\hat{\theta}$  and an innovation sequence obtained by drawing (with replacement) from the set of estimated standardized residuals  $\{\hat{z}_t\}$ . To eliminate the effect of initial values, a length of  $T + \ell$  is usually used, and the first  $\ell$  observations are discarded. In our empirical analysis below we use  $\ell = T$ .

**Step 2:** Fit a normal-ARMA( $p, q$ )-GARCH( $r, s$ ) using the simulated time series  $\{r_t^{(b')}\}$ , and obtain the QML parameter vector estimate  $\hat{\theta}^{(b')}$ .

**Step 3:** Compute a resampled VaR estimate,  $\hat{\kappa}_\lambda^{(b)}(h, T)$ , using the *original series*  $\{r_t\}$ , and the

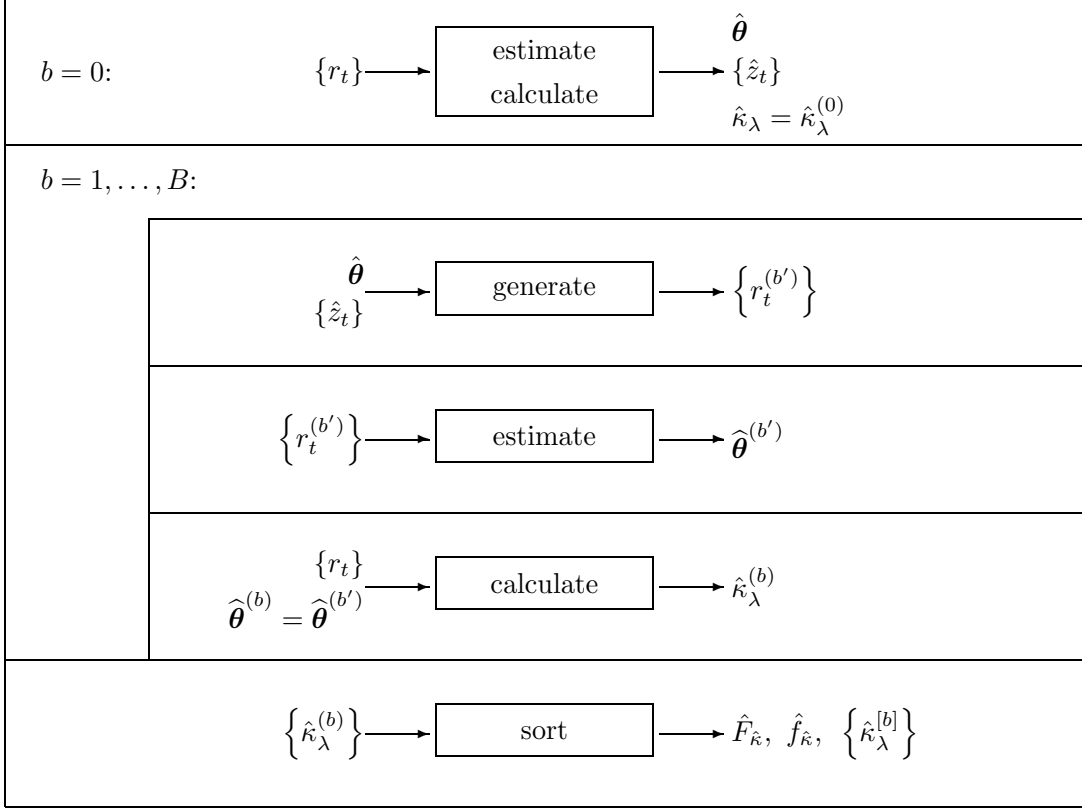


Figure 1: Bootstrap algorithm to obtain distribution of VaR forecasts

bootstrap parameter vector estimate  $\hat{\theta}^{(b)} = \hat{\theta}^{(b')}$ , that is, calculate

$$\hat{\epsilon}_t^{(b)} = r_t - \hat{a}_0^{(b)} - \left( \sum_{i=1}^p \hat{a}_i^{(b)} r_{t-i} \right) - \left( \sum_{j=1}^q \hat{b}_j^{(b)} \hat{\epsilon}_{t-j}^{(b)} \right),$$

$$\hat{\sigma}_t^{2(b)} = \hat{c}_0^{(b)} + \sum_{i=1}^r \hat{c}_i^{(b)} \hat{\epsilon}_{t-i}^{2(b)} + \sum_{j=1}^s \hat{d}_j^{(b)} \hat{\sigma}_{t-j}^{2(b)}, \quad t = 1, \dots, T,$$

to obtain the bootstrapped  $h$ -step-ahead VaR forecast

$$\hat{\kappa}_\lambda^{(b)}(h, T) = \Phi^{-1}(\lambda, \hat{\mu}_{T+h}^{(b)}, \hat{\sigma}_{T+h}^{2(b)}),$$

where

$$\hat{\mu}_{T+h}^{(b)} = \hat{a}_0^{(b)} + \sum_{i=1}^p \hat{a}_i^{(b)} r_{T+h-i} + \sum_{j=1}^q \hat{b}_j^{(b)} \hat{\epsilon}_{T+h-j}^{(b)},$$

$$\hat{\sigma}_{T+h}^{2(b)} = \hat{c}_0^{(b)} + \sum_{i=1}^r \hat{c}_i^{(b)} \hat{\epsilon}_{T+h-i}^{2(b)} + \sum_{j=1}^s \hat{d}_j^{(b)} \hat{\sigma}_{T+h-j}^{2(b)},$$

are calculated in a similar way as in (5), using conditional expectations for unobserved values.

The original VaR prediction  $\hat{\kappa}_\lambda(h, T) =: \hat{\kappa}_\lambda^{(0)}(h, T)$  and the  $B$  bootstrapped VaR predictions,  $\hat{\kappa}_\lambda^{(b)}(h, T)$ ,  $b = 1, \dots, B$ , can now be used to form an empirical distribution of the point estimator

$\hat{\kappa}_\lambda$ , with the empirical distribution function given by

$$\hat{F}_{\hat{\kappa}}(x; \lambda, h, T) = \frac{1}{B+1} \sum_{b=0}^B \mathcal{I}_{(-\infty, x)} \left( \hat{\kappa}_\lambda^{(b)}(h, T) \right), \quad (6)$$

where  $\mathcal{I}$  is the indicator function. Similarly, the kernel density estimate of the  $\hat{\kappa}_\lambda^{(b)}$  will be denoted as  $\hat{f}_{\hat{\kappa}}(\cdot; \lambda, h, T)$ .

Observe that (6) could be used to construct a bootstrap confidence interval for  $\kappa_\lambda$  (which could, for example, be used for delivering a more conservative VaR estimate which takes the estimated parameter uncertainty into account). A related method for doing this has been detailed in Bams, Lehnert and Wolff (2005), and those authors find that, the more complicated the VaR prediction model is, the higher is the uncertainty in the VaR estimate. Our goal is to use resampling in conjunction with a further step to remove the bias in the point estimator of  $\hat{\kappa}_\lambda$ .

## 2.2 Bias-Correcting Value-at-Risk Forecasts

The main idea behind this step is to use past VaR distributions generated by the bootstrap algorithm to adjust the standard (and usually highly biased), VaR predictions of the normal-GARCH model. In light of the definition of VaR, the obvious criterion to use for adjustment is the observed frequency of *violations* (also referred to as *exceptions*), or past realized returns that are less than or equal to the predicted VaR.

For a given probability level  $\lambda$ , the observed frequency of violations, denoted  $\hat{\lambda}$ , for a set of successive VaR predictions obtained from the usual normal-GARCH model between times, say,  $\eta_1$  and  $\eta_2$ , and the corresponding realized returns is given by

$$\hat{\lambda} = \frac{1}{\eta_2 - \eta_1 + 1} \sum_{t=\eta_1}^{\eta_2} \mathcal{I}_{(-\infty, \hat{\kappa}(t)]} (r_{t+h}), \quad \hat{\kappa}(t) = \hat{\kappa}_\lambda(h, t). \quad (7)$$

The observed frequency is less (higher) than the actual risk level  $\lambda$ , if the VaR predictions obtained by the normal-GARCH model tend to overestimate (underestimate) the risk. The aim of our VaR bias-correction is to locate that quantile of past VaR distributions which leads to observed frequencies of VaR violations that coincide (as close as possible) with the given risk level.

Denote by  $\{\hat{\kappa}_\lambda^{[b]}(h, t)\}_{b=0}^B$  the sorted,  $(B+1)$ -length sequence of estimated one-step ahead cut-off points for probability level  $\lambda$  at time-point  $t$  generated by the resampling algorithm with the original normal-GARCH forecast,  $\hat{\kappa}_\lambda^{(0)}(h, t)$ , added, i.e.,  $\hat{\kappa}_\lambda^{[b]}(h, t) \leq \hat{\kappa}_\lambda^{[b+1]}(h, t)$  for  $b = 0, \dots, B-1$ . Finding the correct quantile of the VaR distribution  $\hat{F}_{\hat{\kappa}}$  is equivalent to determining the largest index  $b$ , denoted  $b^*$ , such that, for the corresponding series  $\{\hat{\kappa}_\lambda^{[b^*]}(h, t)\}_{t=\eta_1}^{\eta_2}$ , the observed frequency of violations is less than or equal to the specified risk level  $\lambda$ . To operationalize this, we need a certain number of past VaR distribution forecasts which precede our actual VaR prediction of interest. We consider a moving window procedure for finding the appropriate quantile of the VaR distribution generated by the resampling algorithm.

Our moving window bias-correction procedure involves a fixed number of, say,  $L$  preceding VaR forecast distributions for calculating the appropriate quantile for the  $h$ -step-ahead prediction for the downfall risk made at time  $T$ . The desired quantile or, more precisely, the corresponding  $b^*$ th order statistic, is determined as

$$b^* = \max\{b : b \in \{0, 1, \dots, B\}\} \text{ s.t. } \frac{1}{L} \sum_{t=T-L-h+1}^{T-h} \mathcal{I}_{(-\infty, \hat{\kappa}^{[b]}(h,t))}(r_{t+h}) \leq \lambda. \quad (8)$$

That is,  $b^*$  determines the greatest quantile of the last  $L$  feasible VaR distributions (where feasible means that only the  $h$ -step-ahead VaR distributions for which a corresponding realized return is observed are included and the observed frequency can be calculated) for which the corresponding series of VaR predictions,  $\{\hat{\kappa}_\lambda^{[b^*]}(h,t)\}$  leads to an observed frequency of VaR violations that is equal to (or just smaller) than the given risk level. The reason for the (slight) overestimation of the risk level is just an artifact of using a finite  $B$  and  $L$ , i.e., a limited number of bootstrap replications to obtain the VaR distribution and a limited number of preceding VaR distributions, respectively. For successive calculations of  $b^*$ , the number of preceding VaR distribution forecasts is a constant, namely  $L$ . This means that the oldest VaR distribution of the previous quantile calculation is discarded while the most recent feasible VaR forecast distribution is added to the bias-correction procedure.

The choice of  $L$  is rather subjective and is based on the tradeoff of bias and variance, resulting from the fact that the data generating process is, most likely, changing over time. One could conduct extensive (and time consuming) simulation studies to find (approximately) an optimal  $L$ , but that would only be valid for a particular segment of a particular data set. More preferably, it would be the case that, across “reasonable” choices of  $L$ , the results would not be very sensitive. Fortunately, this is precisely the case, as detailed below in the empirical sections. Briefly, we use  $L = 250$  and  $L = 500$  (corresponding to one and two years of trading data, respectively), and find the results to be not only excellent with regard to several common tests, but, importantly, not particularly sensitive to the choice of  $L$ .

### 3 Empirical Analysis

For testing our method, we use the three major stock indices DAX, NASDAQ Composite, and the Nikkei 225, and the two currency exchange rates Japanese-Yen/US-Dollar and the US-Dollar/British Pound. The starting date for the series is February 06, 1991 for the DAX; February 14, 1991 for the NASDAQ; October 15, 1990 for the Nikkei; and February 01, 1991 for the Yen and Pound, respectively. In addition, we use simulated data, so that the performance of the method under a known data generating process can be examined. For both real and simulated data, we will see that the method performs very well.

From the restricted normal-GARCH class (1) and (2), the usual values of  $p = r = s = 1$  and  $q = 0$  were found to be adequate for capturing the dynamics in the mean and variance. The estimation period is set to  $T = 1,000$ , which corresponds to about four years of daily returns

data for estimating the parameters of the model. For calculating the VaR forecast distributions,  $B = 500$  bootstrap replications are used.

For each series, we work with  $h = 1$  step-ahead VaR predictions, using  $P = 2,000$  out-of-sample values, based on the downfall probabilities  $\lambda = 0.01, 0.02, \dots, 0.10$ , where the last forecast is made for December 31, 2004 for all the series. For measuring the effect of different window lengths on the bias-correction method, we use the two values  $L = 250$  and  $L = 500$ . This implies an additional number of  $L = 500$  VaR forecast distributions resulting in a total of  $P + L = 2,500$  VaR forecast distributions. So, the date for the first forecast distribution needed in our application is February 02, 1995 for the DAX; January 30, 1995 for the NASDAQ; November 11, 1994 for the Nikkei; and January 24, 1995 for the Yen and Pound, respectively.

We use two competing models to assess the quality of our bias-correction procedure. The first is just the normal-AR(1)-GARCH(1,1) model, which is a by-product of calculating the VaR forecast distributions, i.e.,  $\hat{\kappa} = \hat{\kappa}^0$ . The second, and more worthy model we use for comparison is the  $t$ -AR(1)-GARCH(1,1) model. The corresponding VaR forecast for the  $t$ -AR(1)-GARCH(1,1) model is given by

$$\hat{\kappa}_\lambda = \hat{\kappa}_\lambda(h, T) = \hat{\mu}_{T+h} + \hat{\sigma}_{T+h} F_t^{-1}(\lambda; \hat{\nu}),$$

where  $F_t^{-1}(\lambda; \nu)$  is the inverse of the cumulative distribution function of the  $t$ -distribution with  $\nu$  degrees of freedom and a standardized variance of one.

In the following, we first inspect the VaR forecasts distributions obtained by the bootstrap algorithm, and then examine how the optimal quantiles for the bias-correction step differ for different window lengths  $L$ . Lastly, and most importantly, we compare the VaR predictions of the competing models.

### 3.1 Value-at-Risk Forecast Distribution

Figures 2 and 3 present exemplary VaR forecast distributions for the DAX and British pound data (the others are similar and available upon request). The upper part of the figures refer to the first VaR forecast distribution included in our analysis, while the lower part of the figures refer to the last VaR forecast distribution, corresponding to 31 December 2004. In addition to the VaR forecast distributions, the vertical lines refer to the usual normal-GARCH VaR prediction,  $\hat{\kappa}^0$ .

Clearly evident from the figures is the increasing risk in VaR predictions for smaller probability levels. The VaR forecast distributions become flatter for smaller probability levels. Another interesting feature readily apparent from the figures is that the usual normal-GARCH VaR prediction does not necessarily coincide with the mode of the VaR forecast distribution, but it is always near its median. For example, for the last VaR forecast made for the DAX, the usual normal-GARCH VaR forecast is smaller than the mode of the VaR forecast distribution. This is also true for the first forecast for NASDAQ, the first forecast for Nikkei, and the first forecast for the British pound. On the other hand, the last VaR forecast for the NASDAQ (and the last forecast for the Japanese yen) from the usual normal-GARCH model are greater than the mode of the corresponding VaR forecast distributions.

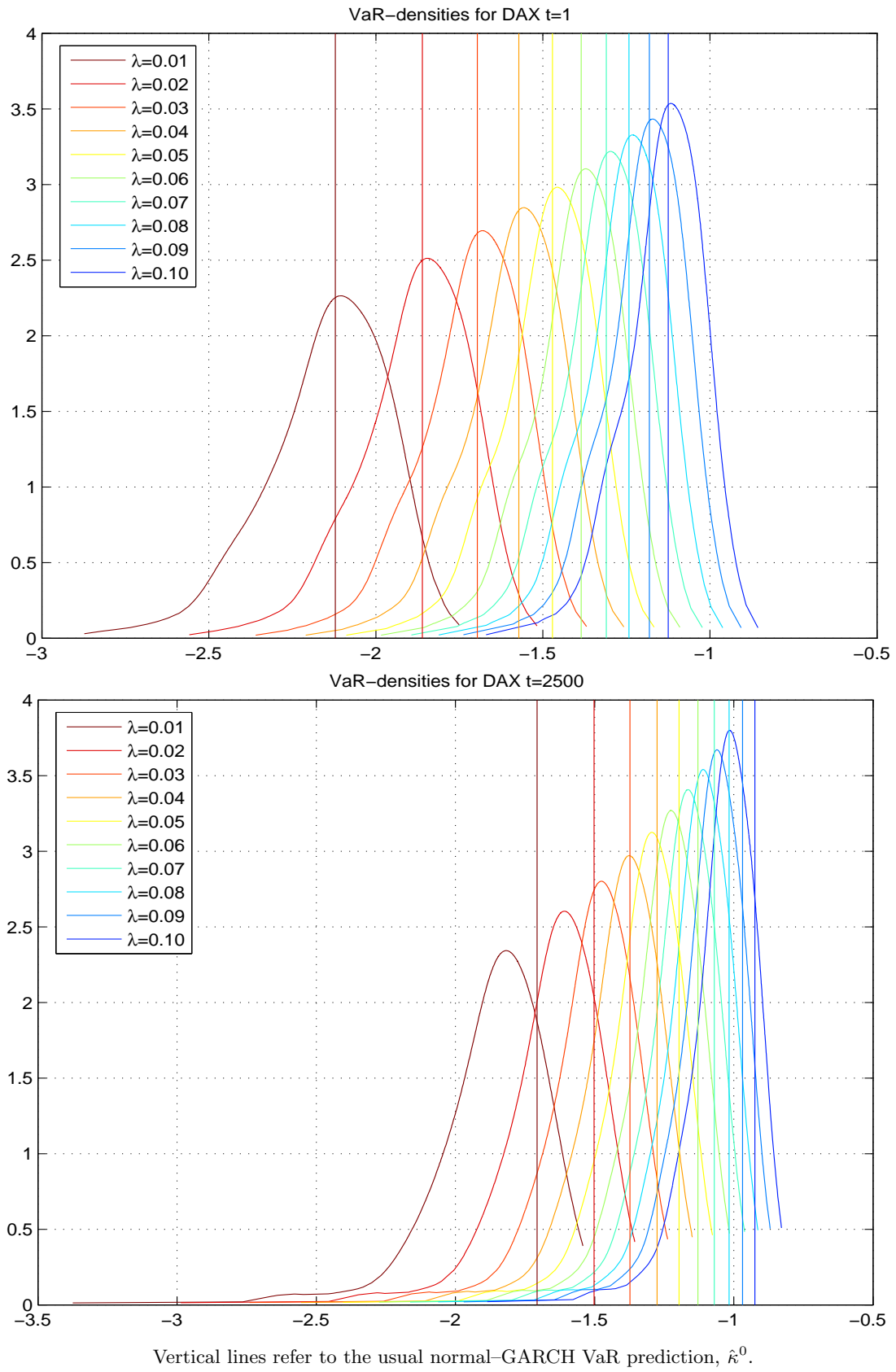


Figure 2: VaR forecast distributions for the first and last forecast - DAX

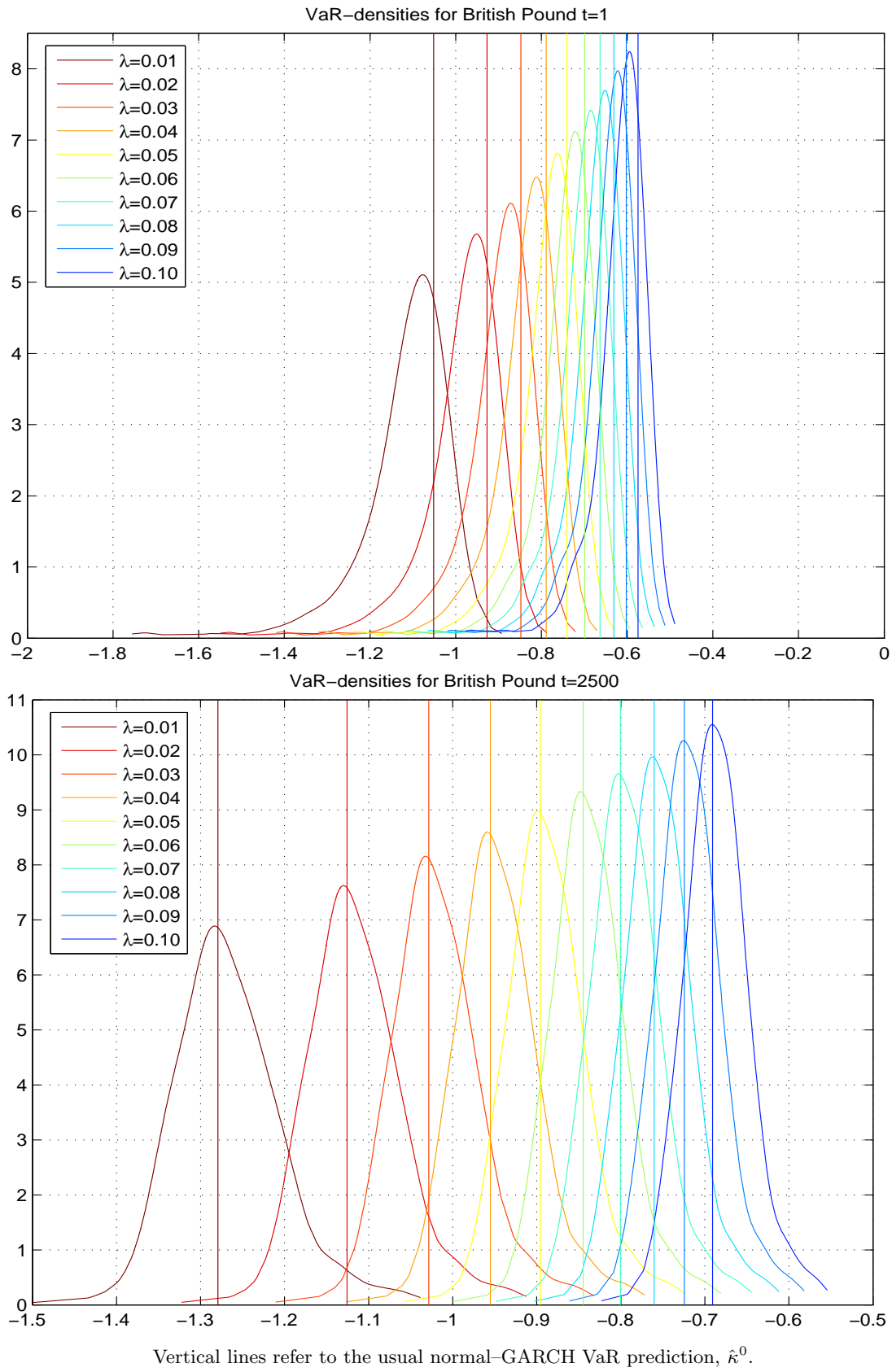


Figure 3: VaR forecast distributions for the first and last forecast - British pound

### 3.2 Bias-Correcting Value-at-Risk Forecasts

To better understand the workings of the bias-correction procedure, Figures 4 and 5 show the optimal quantiles,  $b^*/(B + 1)$ , for the downfall probabilities  $\lambda = 0.01, 0.05, 0.10$  over the forecasting period for the DAX and British pound data, respectively (those for the other series and downfall probabilities are similar, and available upon request). The upper (lower) part of the figures refers to a window of length  $L = 250$  ( $L = 500$ ).

The resulting optimal prediction quantiles for both window lengths look similar, but do not fully coincide. Usually the means of the optimal quantiles for both window lengths are close to one another. While the smaller window implies a higher influence of individual VaR violations within the window, the variance of the optimal quantiles is higher for the shorter window (tabulated details of this are available upon request).

We see that the optimal prediction quantiles move along all possible values, i.e., there can be large fluctuations of the optimal quantile. This implies that, while the usual normal-GARCH VaR prediction is near the median of the VaR forecast distribution, it is obvious that there are times for which this usual VaR prediction is overestimating the downfall probability, as well as times which it is underestimating.

The correlation between the quantiles of different downfall probabilities decreases as the difference between the compared downfall probabilities increases (full details are available upon request). Thus, an individual adjustment for each downfall probability, as proposed, seems necessary. Of course, the reported quantiles are chosen to be optimal for the past realized violations. Testing these optimal quantiles for genuine out-of-sample VaR forecasts is necessary, and now considered. For testing the accuracy of the different models for predicting VaR, we follow the detailed textbook description in Christoffersen (2003). For  $h = 1$  step-ahead VaR predictions,  $\hat{\kappa}_\lambda(1, t)$ , and observed actual returns  $r_{t+1}$ , the Boolean sequence indicating the presence or absence of VaR violations is defined as

$$I_{t+1} = \mathcal{I}_{(-\infty, \hat{\kappa}_\lambda(1, t))}(r_{t+1}).$$

With  $T_1 = \sum_{t=1}^T I_{t+1}$  the number of violations, and  $T_0 = T - T_1$  the number of non-violations, the empirical downfall probability is given by  $\hat{\lambda} = T^{-1} \sum_{t=1}^T I_{t+1} = T_1/T$ . For a correct VaR prediction model, we expect the violation sequence  $I_{t+1}$  to be

$$H_0 : I_{t+1} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(\lambda).$$

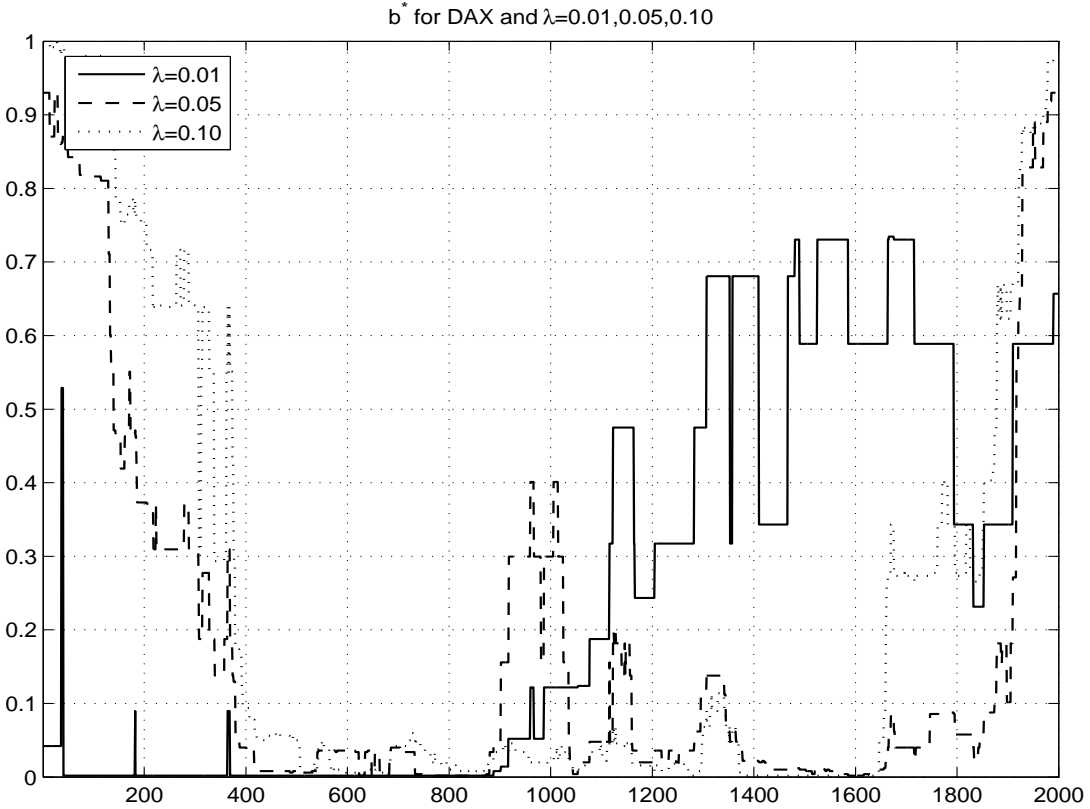
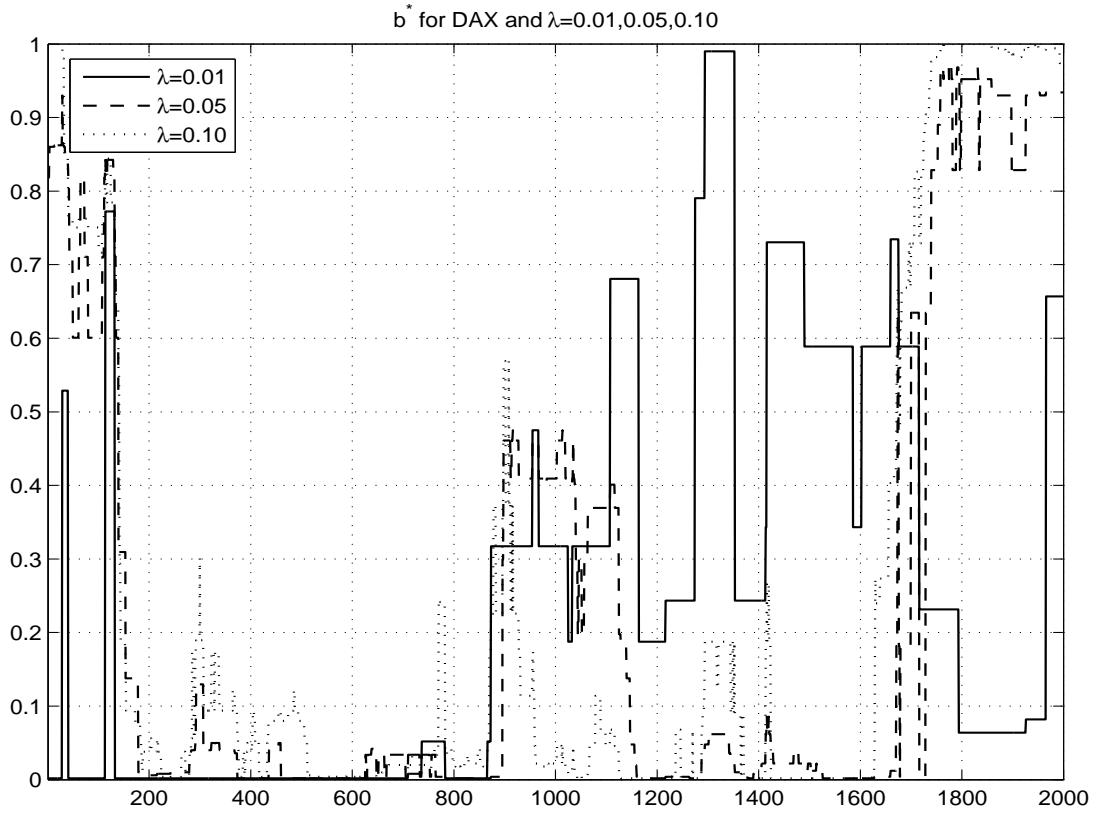
Testing this null hypothesis is twofold. One part is testing the unconditional coverage, or that the observed downfall probability is equal to the specified downfall probability (unconditional coverage). The second part is the iid-ness of the violations (independence).

For the first part, the likelihood value under the null hypothesis that  $\hat{\lambda} = \lambda$  is

$$L(\lambda) = \prod_{t=1}^T (1 - \lambda)^{1 - I_{t+1}} \lambda^{I_{t+1}} = (1 - \lambda)^{T_0} \lambda^{T_1},$$

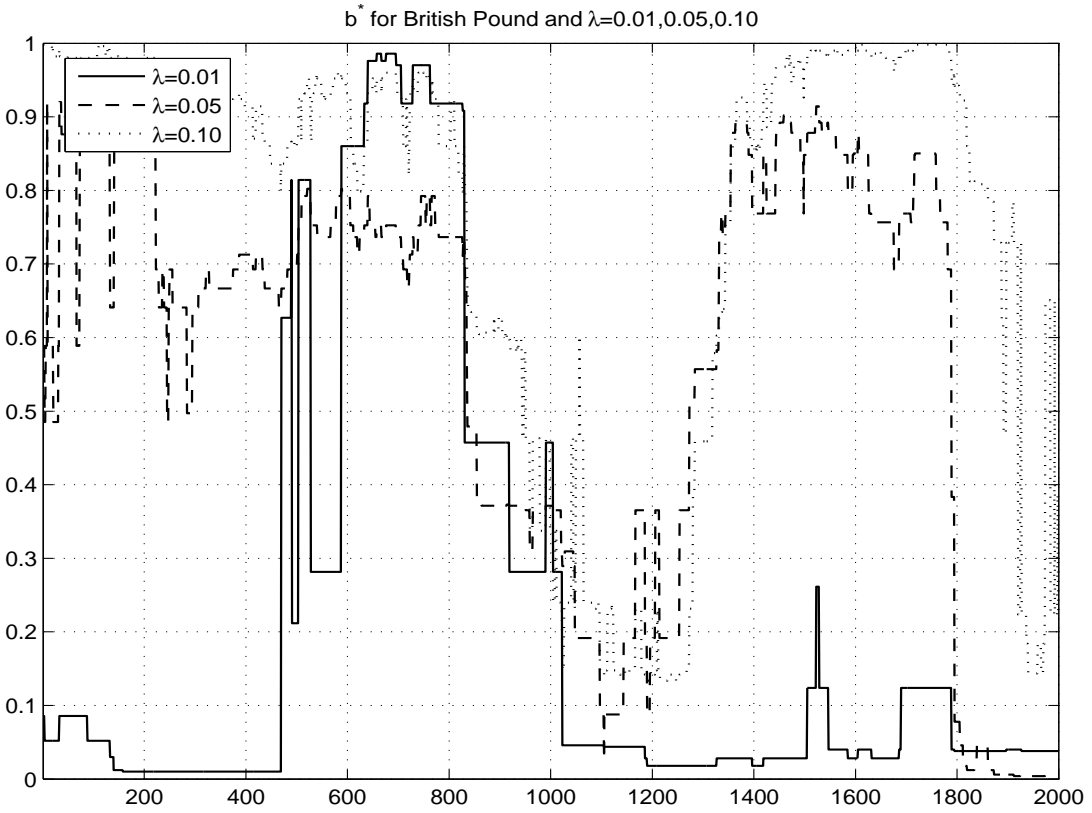
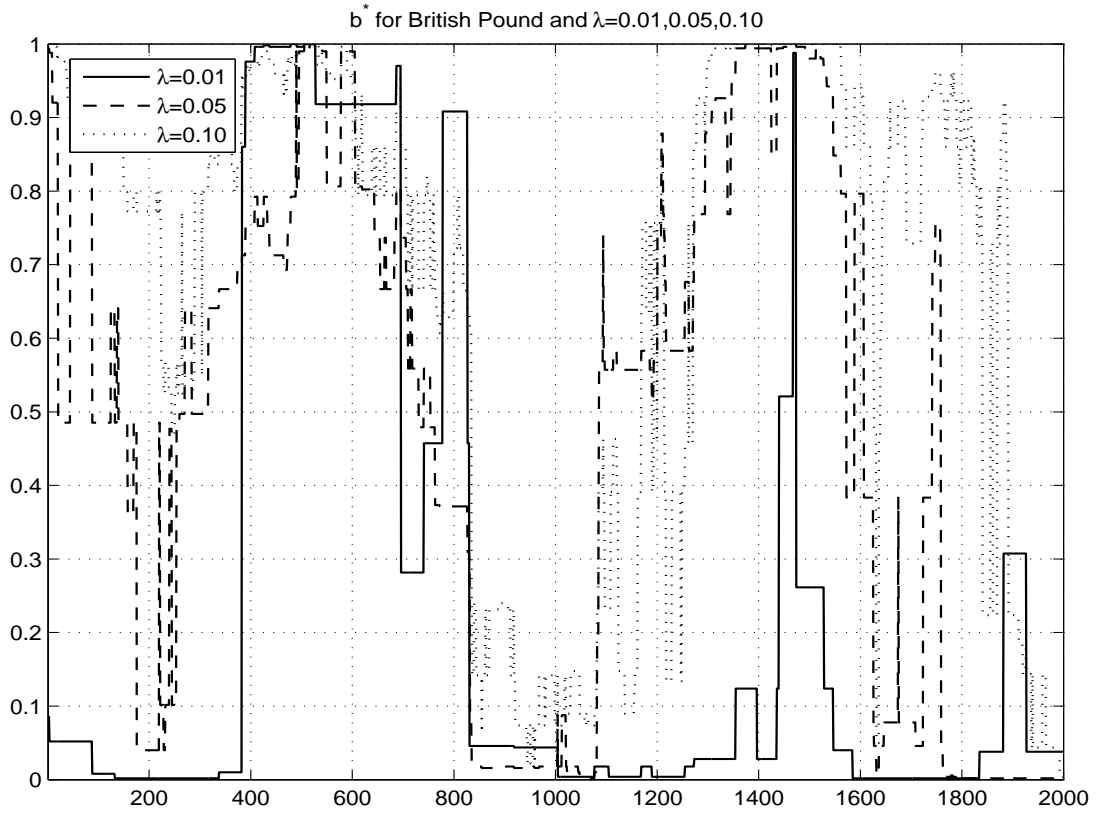
while the observed likelihood value is given by  $L(\hat{\lambda}) = (1 - \hat{\lambda})^{T_0} \hat{\lambda}^{T_1}$ . The unconditional coverage is tested by using the likelihood ratio test statistic and the corresponding, asymptotically valid





Upper (lower) figure refers to  $L = 250$  ( $L = 500$ )

Figure 4: Optimal quantile  $b^*/(B + 1)$  for VaR prediction - DAX



Upper (lower) figure refers to  $L = 250$  ( $L = 500$ )

Figure 5: Optimal quantile  $b^*/(B + 1)$  for VaR prediction - British pound

$p$ -value

$$\text{LR}_{\text{uc}} = -2 \ln[L(\lambda)/L(\hat{\lambda})] \sim \chi_1^2, \quad \text{P}_{\text{uc}} = 1 - F_{\chi_1^2}(\text{LR}_{\text{uc}}),$$

i.e.,  $\text{P}_{\text{uc}}$  is the probability of getting a sample that conforms even less to the null hypothesis than the sample. For  $\text{P}_{\text{uc}}$  below the desired significance level, the null hypothesis is rejected.

For testing the independence of  $I_{t+1}$ , as in Christoffersen (1998), let  $\Lambda$  be the transition probability matrix for a first order Markov sequence,

$$\Lambda = \begin{bmatrix} \lambda_{00} & \lambda_{01} \\ \lambda_{10} & \lambda_{11} \end{bmatrix},$$

where  $\lambda_{ij}$  are the proportions given by  $\lambda_{ij} = \text{prop}(I_t = i \text{ and } I_{t+1} = j)$ ,  $i, j = 0, 1$ . With  $T_{ij}$ ,  $i, j = 0, 1$ , the number of observations with a  $j$  following an  $i$ , the observed probabilities are given by

$$\hat{\lambda}_{01} = \frac{T_{01}}{T_{00} + T_{01}}, \quad \hat{\lambda}_{11} = \frac{T_{11}}{T_{10} + T_{11}},$$

and  $\hat{\lambda}_{00} = 1 - \hat{\lambda}_{01}$ ,  $\hat{\lambda}_{10} = 1 - \hat{\lambda}_{11}$ . The likelihood value under the null, i.e.,  $\hat{\lambda}_{01} = \hat{\lambda}_{11} = \hat{\lambda}$ , is  $L(\hat{\lambda}) = (1 - \hat{\lambda})^{T_0} \hat{\lambda}^{T_1}$  and the observed likelihood value is given by

$$L(\hat{\Lambda}) = (1 - \hat{\lambda}_{01})^{T_{00}} \hat{\lambda}_{01}^{T_{01}} (1 - \hat{\lambda}_{11})^{T_{10}} \hat{\lambda}_{11}^{T_{11}}.$$

Again, a likelihood ratio test statistic and corresponding  $p$ -value

$$\text{LR}_{\text{ind}} = -2 \ln[L(\hat{\lambda})/L(\hat{\Lambda})] \sim \chi_1^2, \quad \text{P}_{\text{ind}} = 1 - F_{\chi_1^2}(\text{LR}_{\text{ind}})$$

are used for testing the independence of the VaR violations.

As in Christoffersen (1998), a test for the conditional coverage involves the two hypotheses  $\hat{\lambda} = \lambda$  and  $I_{t+1}$  are iid, and the unconditional coverage and the independence test statistics can be combined. The likelihood ratio test statistic with corresponding  $p$ -value is given by

$$\text{LR}_{\text{cc}} = -2 \ln[L(\lambda)/L(\hat{\Lambda})] = \text{LR}_{\text{uc}} + \text{LR}_{\text{ind}} \sim \chi_2^2, \quad \text{P}_{\text{cc}} = 1 - F_{\chi_2^2}(\text{LR}_{\text{cc}}).$$

Our comparison is based on the asymptotically valid  $p$ -values for three test statistics described above, and models with higher  $p$ -values are preferred.

In addition to the coverage test statistics we also report the mean relative scaled bias (MRSB) for the different models used for forecasting VaR as described in Engel and Gizycki (1999). The MRSB determines which model produces the smallest average VaR predictions, when VaR forecasts are suitable scaled to obtain the desired downfall risk.

The calculation of the MRSB consists of two steps. The first step is to calculate *ex-post* multipliers,  $X_{i,\lambda}$  for the  $N$  different models,  $i = 1, \dots, N$  and downfall risks  $\lambda$ , which are needed to obtain the correct unconditional coverage. The number of violations for model  $i$  and downfall risk  $\lambda$ ,  $T_{1,i,\lambda}$ , after scaling must coincide with the expected number of downfalls, i.e., choose  $X_{i,\lambda}$  so that

$$T_{1,i,\lambda} = \lambda T, \quad \text{with} \quad T_{1,i,\lambda} = \sum_{t=1}^T \mathcal{I}_{(-\infty, \hat{\kappa}_{i,\lambda}(1,t) \cdot X_{i,\lambda})}(r_{t+1}).$$

In the second step, the scaled VaR predictions,  $Y_{i,\lambda,t} = \hat{\kappa}_{i,\lambda}(1,t) \cdot X_{i,\lambda}$ , are used to calculate the mean relative scaled bias of model  $i$  for downfall risk  $\lambda$  as

$$\text{MRSB}_{i,\lambda} = 100 \times \frac{1}{T} \sum_{t=1}^T \frac{Y_{i,\lambda,t} - \bar{Y}_{\lambda,t}}{\bar{Y}_{\lambda,t}}, \quad \text{where} \quad \bar{Y}_{\lambda,t} = \frac{1}{N} \sum_{i=1}^N Y_{i,\lambda,t}.$$

When interpreting the MRSB results, it must be kept in mind that ex-post multipliers for obtaining correct coverage of the models' VaR predictions are used, i.e., the MRSB is a measure which is *conditional on correct coverage*. In addition, the MRSB is a relative measure, in that it considers the deviation from the mean of the different VaR predictions. So, they depend on the choice and number of models used in the comparison, and the results would change by adding or removing models from the forecasting comparison.

The results based on MRSB are not in favor of any model and the resulting values are relatively small. In light of the above comments on MRSB and keeping in mind that we compare only four models, the mean VaR prediction should be very noisy. As such, this measure of efficiency is not particularly useful in this comparison and, in the following, we concentrate on the measures  $\text{LR}_{\text{uc}}$ ,  $\text{LR}_{\text{ind}}$  and  $\text{LR}_{\text{cc}}$ .

The results for the four competing models are given in Tables 1 and 2 for the DAX and British pound (the others are similar and available upon request).

We first focus on the ability of our procedure to improve the VaR forecasts of the normal-GARCH model. Comparing the forecasts from the latter model with the bias-corrected forecasts for  $L = 250$ , we find, for all series except the British pound, increasing  $P_{\text{uc}}$ -values for at least 7 of the 10 specified downfall probabilities. For the DAX and the NASDAQ, we have increasing  $P_{\text{uc}}$ -values for all proposed probabilities. Only for the British pound series did we find half of the test statistics in favor of the usual forecasts and the other half preferring the bias-corrected VaR forecasts.

More interestingly, we found no  $P_{\text{uc}}$ -values below the 10% significance level for the bias-corrected VaR predictions with  $L = 250$  (and only 2 (1) below the 10% (5%) level for  $L = 500$ ), while for the usual VaR forecasts, we have at least 2  $p$ -values below the 10% level, with a maximum number of 10 observed frequencies significantly different from the proposed risk level for the DAX.

Summarizing the results for the unconditional coverage, our method is able to improve the normal-GARCH VaR forecasts so much that VaR predictions are obtained which are insignificantly different from the proposed downfall probability. Observe that this pertains to a genuine out-of-sample forecast exercise, and does not involve using the results from the “training sample” used for estimation and bias correction.

Regarding independence of the VaR violations over time, for both the usual normal-GARCH VaR forecast violations and those from the new bias-correction method, independence cannot be rejected, i.e., the occurrence of violations for usual VaR forecasts have no systematic pattern. The new method yields a small increase of significant test statistics for both bias-correction window lengths.

Combining the two aforementioned test statistics and looking at the conditional coverage

Table 1: Unconditional Coverage, Independence, Conditional Coverage - DAX

	$\lambda$	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900	0.1000
$\mathbb{P}^c$	$\hat{\lambda}_N$	0.0160	0.0280	0.0435	0.0565	0.0675	0.0790	0.0905	0.0980	0.1080	0.1230
	$\hat{\lambda}_t$	0.0125	0.0270	0.0415	0.0565	0.0680	0.0820	0.0950	0.1030	0.1150	0.1290
	$\hat{\lambda}_{250}^*$	0.0100	0.0250	0.0335	0.0450	0.0535	0.0660	0.0745	0.0875	0.0930	0.1040
	$\hat{\lambda}_{500}^*$	0.0110	0.0245	0.0330	0.0430	0.0540	0.0660	0.0745	0.0840	0.0950	0.1030
$\mathbb{P}^d$	$\hat{\lambda}_N$	0.0131	0.0159	0.0009	0.0004	0.0006	0.0006	0.0006	0.0040	0.0062	0.0009
	$\hat{\lambda}_t$	0.2794	0.0337	0.0043	0.0004	0.0004	0.0001	0.0000	0.0003	0.0002	0.0000
	$\hat{\lambda}_{250}^*$	1.0000	0.1240	0.3676	0.2630	0.4774	0.2657	0.4347	0.2227	0.6409	0.5533
	$\hat{\lambda}_{500}^*$	0.6582	0.1648	0.4388	0.4986	0.4175	0.2657	0.4347	0.5128	0.4384	0.6561
$\mathbb{P}^e$	$\hat{\lambda}_N$	0.3002	0.6777	0.2760	0.7051	0.3871	0.0627	0.1884	0.2336	0.1718	0.1522
	$\hat{\lambda}_t$	0.4171	0.6200	0.3554	0.7051	0.5574	0.3922	0.2385	0.2497	0.2635	0.3968
	$\hat{\lambda}_{250}^*$	0.5148	0.7433	0.7543	0.2251	0.0566	0.2711	0.4205	0.2932	0.2062	0.1383
	$\hat{\lambda}_{500}^*$	0.4744	0.7680	0.7729	0.6230	0.3629	0.2711	0.5920	0.5943	0.3527	0.1640
MRSB	$\hat{\lambda}_N$	0.0270	0.0501	0.0022	0.0017	0.0020	0.0005	0.0011	0.0079	0.0093	0.0014
	$\hat{\lambda}_t$	0.4009	0.0927	0.0111	0.0017	0.0018	0.0003	0.0001	0.0007	0.0005	0.0001
	$\hat{\lambda}_{250}^*$	0.8088	0.2904	0.6345	0.2561	0.1262	0.2938	0.5329	0.2736	0.4034	0.2797
	$\hat{\lambda}_{500}^*$	0.7021	0.3648	0.7108	0.7048	0.4759	0.2938	0.6384	0.7005	0.4808	0.3439
MRSB	$\hat{\lambda}_N$	0.2212	0.4331	-1.8172	-1.1841	-1.2201	-0.4710	0.0264	-0.2757	0.2147	-0.8273
	$\hat{\lambda}_t$	-0.5738	0.2141	-1.3706	-1.2210	-1.3480	-0.9505	-0.7250	-0.1778	1.0802	0.3325
	$\hat{\lambda}_{250}^*$	-0.5670	-0.8677	2.3441	1.9202	1.6961	1.8669	1.7192	0.2843	-0.7384	0.6506
	$\hat{\lambda}_{500}^*$	0.9196	0.2206	0.8436	0.4849	0.8720	-0.4454	-1.0206	0.1693	-0.5565	-0.1559

$\hat{\lambda}_N$ : observed downfall probability for normal-AR(1)-GARCH(1,1) model

$\hat{\lambda}_t$ : observed downfall probability for  $t$ -AR(1)-GARCH(1,1) model

$\hat{\lambda}_{250}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 250$

$\hat{\lambda}_{500}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 500$

For all models:  $T = 1000$ , and  $P = 2000$  Value-at-Risk forecasts included

$\mathbb{P}^c$  ( $\mathbb{P}^d$ ,  $\mathbb{P}^e$ ): Probability of observing a sample with higher unconditional coverage test

statistic (independence test statistic, conditional coverage test statistic).

MRSB: Mean Relative Scaled Bias

\*/\*\*/\*\*: number of corresponding  $p$ -values smaller than 0.10/0.05/0.01

Table 2: Unconditional Coverage, Independence, Conditional Coverage - British pound

$\lambda$	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900	0.1000
$\hat{\lambda}_N$	0.0135	0.0280	0.0365	0.0425	0.0485	0.0635	0.0745	0.0805	0.0865	0.0920
$\hat{\lambda}_t$	0.0075	0.0225	0.0340	0.0450	0.0565	0.0705	0.0815	0.0900	0.1010	0.1075
$\hat{\lambda}_{250}^*$	0.0110	0.0225	0.0335	0.0460	0.0545	0.0675	0.0740	0.0840	0.0950	0.1050
$\hat{\lambda}_{500}^*$	0.0135	0.0240	0.0350	0.0440	0.0555	0.0670	0.0770	0.0860	0.0990	*/**/***/
$\hat{\lambda}_N$	0.1353	0.0159	0.0991	0.5721	0.7571	0.5136	0.4347	0.9344	0.5822	0.2273
$\hat{\lambda}_t$	0.2397	0.4336	0.3043	0.2630	0.1909	0.0540	0.0491	0.1055	0.0912	0.2687
$\hat{\lambda}_{250}^*$	0.6582	0.4336	0.3676	0.1808	0.3624	0.1657	0.4870	0.5128	0.4384	0.4593
$\hat{\lambda}_{500}^*$	0.1353	0.2153	0.2013	0.3687	0.2671	0.1952	0.2267	0.3280	0.1657	0.1854
$\hat{\lambda}_N$	0.3813	0.5803	0.1732	0.4350	0.7298	0.1506	0.1201	0.0764	0.1527	0.1061
$\hat{\lambda}_t$	0.6229	0.1455	0.2803	0.3179	0.1463	0.0957	0.0533	0.1151	0.1034	0.1580
$\hat{\lambda}_{250}^*$	0.4744	0.1455	0.0299	0.6251	0.5764	0.1738	0.0591	0.0932	0.1209	0.2161
$\hat{\lambda}_{500}^*$	0.3813	0.1205	0.6588	0.7586	0.2342	0.0430	0.0331	0.1390	0.1080	0.2052
$\hat{\lambda}_N$	0.2235	0.0468	0.1015	0.6286	0.8981	0.2875	0.2203	0.2074	0.3091	0.1308
$\hat{\lambda}_t$	0.4440	0.2552	0.3294	0.3245	0.1480	0.0390	0.0223	0.0780	0.0638	0.2002
$\hat{\lambda}_{250}^*$	0.7021	0.2552	0.0630	0.3624	0.5651	0.1518	0.1322	0.1973	0.2224	0.3538
$\hat{\lambda}_{500}^*$	0.2235	0.1390	0.4010	0.6369	0.2662	0.0557	0.0497	0.2075	0.1052	0.1865
$\hat{\lambda}_N$	-0.7798	-1.4036	-1.3169	-0.3045	-3.1036	0.1687	1.3223	-0.8585	-2.0124	-0.7584
$\hat{\lambda}_t$	-0.9203	-0.4516	-0.3880	-2.5851	0.7200	-0.5495	-0.1698	0.3266	-0.0868	0.6264
$\hat{\lambda}_{250}^*$	1.2233	-0.7466	0.2676	2.3212	2.6103	0.1004	-0.8565	0.2044	0.3371	0.3234
$\hat{\lambda}_{500}^*$	0.4768	2.6018	1.4373	0.5684	-0.2266	0.2804	-0.2960	0.3275	1.7621	-0.1914

$\hat{\lambda}_N$ : observed downfall probability for normal-AR(1)-GARCH(1,1) model

$\hat{\lambda}_t$ : observed downfall probability for  $t$ -AR(1)-GARCH(1,1) model

$\hat{\lambda}_{250}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 250$

$\hat{\lambda}_{500}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 500$

For all models:  $T = 1000$ , and  $P = 2000$  Value-at-Risk forecasts included

$P_{uc}$  ( $P_{ind}$ ,  $P_{cc}$ ): Probability of observing a sample with higher unconditional coverage test

statistic (independence test statistic, conditional coverage test statistic).

MRSB: Mean Relative Scaled Bias

\*/\*\*/\*\*\*/: number of corresponding  $p$ -values smaller than 0.10/0.05/0.01

test statistic, we see that, for all but the British pound series, the  $p$ -values for at least 8 out of the 10 risk probabilities for  $L = 250$  (and at least 7 out of 10 for  $L = 500$ ) are superior (larger) than those for the usual normal-GARCH model. Only for the British pound are the results not that clear. For  $L = 250$ , the usual forecasts result in higher  $p$ -values for 7 risk probabilities, while for  $L = 500$ , the usual forecasts are preferred in 4 out of the ten cases.

With respect to the different window lengths used for bias-correction, the results slightly prefer the shorter window of length  $L = 250$  with a higher number of greater  $p$ -values and a smaller number of significant test statistics. As this holds for all considered time series, it appears safe to say that the method is not overly sensitive to the choice of  $L$ .

The promising results obtained by using bias-corrected VaR forecast instead of the usual normal-GARCH forecasts carry over to the VaR forecast results of the  $t$ -GARCH model. The  $t$ -GARCH distribution results for the DAX and NASDAQ data sets yield a large number of significant unconditional coverage test statistics, thus indicating the inability of this model to correctly forecast the VaR. While the independence test statistics indicates no systematic problems, the conditional coverage test statistic carries over the bad results from the unconditional coverage test statistic. For only one downfall probability for the Nikkei index, one for the Japanese Yen series, and one for the British pound series, the  $t$ -GARCH VaR forecasts are superior to the normal-GARCH forecasts and the bias-corrected GARCH VaR forecasts based on a bias-correction window length of  $L = 250$  and  $L = 500$ , while they are never better for DAX and NASDAQ.

### 3.3 Simulation Results

While the usefulness of any method is best judged using a variety of real data sets, as was considered above, it is nevertheless valuable to use simulated data, so as to gain an understanding of the capability of the model under realistic—and controllable—circumstances. To this end, we generate time series from an asymmetric power ARCH, or A-PARCH model, introduced by Ding et al. (1993), which, in comparison to the normal-GARCH model, allows the power term of the volatility equation to be different from two, and, more importantly, also captures asymmetric effects of return shocks on volatility. To make the data generating process even more realistic, we assume the innovations of the A-PARCH model to be Student's  $t$ -distributed, the combination of which has been shown by Mittnik and Paolella (2000) and Giot and Laurent (2004) to be a very competitive model for fitting asset returns.

The  $t$ -A-PARCH model for returns  $r_t$  is given by

$$r_t = a_0 + \sum_{i=1}^p a_i r_{t-i} + \epsilon_t + \sum_{j=1}^q b_j \epsilon_{t-j}, \quad \epsilon_t = z_t \sigma_t, \quad z_t \sim t_\nu, \quad (9)$$

$$\sigma_t^\delta = c_0 + \sum_{i=1}^r c_i (|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^\delta + \sum_{j=1}^s d_j \sigma_{t-j}^\delta. \quad (10)$$

We generate 10 samples from the AR(1)-A-PARCH(1,1) model using the following parameter

values, which are typical for asset return series:

$$\begin{aligned}
 a_0 &= 0.050, & a_1 &= 0.200, \\
 c_0 &= 0.035, & c_1 &= 0.200, & d_1 &= 0.700, \\
 \gamma_1 &= -0.200, & \delta &= 1.600, & \nu &= 5.000.
 \end{aligned}$$

To avoid any starting-value problems, a “burn in” segment was used, i.e., the first 500 observations of each series was discarded. The same forecasting exercise as used on the above empirical study was conducted on the simulated series. We constructed  $P = 2,000$  out-of-sample  $h = 1$  step-ahead VaR predictions based on downfall probabilities  $\lambda = 0.01, 0.02, \dots, 0.10$  with an estimation period length  $T = 1,000$ , and  $B = 500$  bootstrap replications. As before, we use the VaR forecasts of a normal-AR(1)-GARCH(1,1) model and the bias-corrected VaR predictions of the normal-AR(1)-GARCH(1,1) model with window lengths  $L = 250$  and  $L = 500$  for the bias-correction.

The detailed results for the first of the 10 simulated series are given in Table 3 (the full set of tables are available on request from the authors). As we are focusing on the  $p$ -values to judge the VaR forecasting ability of a model, we report the number of  $p$ -values below the desired significance level for all the simulated samples in Table 4. We see that, for all samples, the usual normal-GARCH VaR predictions fail to predict the downfall risk accurately for the range of downfall probabilities chosen. This is not surprising, given that the data generating process is far more realistic, and differs markedly from the normal-GARCH model. In stark comparison, the bias-correction method yields quite high accuracy, for both window lengths.

For all samples, the number of  $P_{uc}$  and  $P_{cc}$  values below the significance levels are highly reduced by our method, regardless of the window length  $L$  chosen, with a small preference for the longer window length  $L = 500$ . For  $P_{ind}$ , the number of  $p$ -values below the significance level are not affected by the bias-correction in any particular direction. This simulation exercise is clearly limited to a single, albeit general and popularly-assumed, data generating process (DGP), and strongly indicates the validity of the method for data which exhibit the usual stylized facts over and above what a normal-GARCH model can capture, such as conditional power tails for the innovations and asymmetric volatility responses. Further simulation could be conducted using other DGPs, such as higher order GARCH processes (i.e.,  $r + s > 2$ ), or fractionally integrated GARCH, or Markov-switching GARCH, or GARCH processes with richer correlation dynamics and time-varying skewness, such as given by the mixed-normal-GARCH model class of Haas, Mittnik and Paolella (2004) and Alexander and Lazar (2006), or stochastic volatility models.

## 4 Conclusions

We present a resampling and bias-correction method based on the normal-GARCH model for delivering easily computed and accurate VaR forecasts. The method is very easy to implement, and is fully data driven. The results are robust to the chosen bias-correction window length, with a slight preference for the smaller window length of  $L = 250$  for the five real return series



Table 3: Unconditional Coverage, Independence, Conditional Coverage - Sample 1

$\lambda$	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900	0.1000
$\hat{\lambda}_N$	0.0115	0.0185	0.0260	0.0335	0.0395	0.0465	0.0550	0.0605	0.0675	0.0715
$\hat{\lambda}_{250}^*$	0.0130	0.0250	0.0305	0.0440	0.0505	0.0605	0.0680	0.0770	0.0880	0.0970
$\hat{\lambda}_{500}^*$	0.0105	0.0205	0.0290	0.0410	0.0495	0.0585	0.0675	0.0740	0.0855	*/**/**
$\hat{\lambda}_N$	0.5102	0.6275	0.2835	0.1273	0.0255	0.0083	0.0064	0.0008	0.0002	0.0000
$\hat{\lambda}_{250}^*$	0.1974	0.1240	0.8960	0.3687	0.9184	0.9251	0.7248	0.6189	0.7538	0.6533
$\hat{\lambda}_{500}^*$	0.8236	0.8736	0.7921	0.8202	0.9182	0.7767	0.6595	0.3170	0.4786	0.2581
$\hat{\lambda}_N$	0.4548	0.6804	0.6902	0.3198	0.4475	0.1814	0.1389	0.3048	0.1040	0.1151
$\hat{\lambda}_{250}^*$	0.3990	0.1058	0.0482	0.2582	0.0962	0.3048	0.5574	0.4706	0.5470	0.2701
$\hat{\lambda}_{500}^*$	0.4944	0.1848	0.0604	0.3773	0.1137	0.1968	0.2188	0.1541	0.2473	0.1352
$\hat{\lambda}_N$	0.6089	0.8166	0.5196	0.1906	0.0619	0.0125	0.0082	0.0022	0.0003	0.0000
$\hat{\lambda}_{250}^*$	0.3054	0.0828	0.1409	0.3524	0.2494	0.5880	0.7913	0.6812	0.7941	0.4922
$\hat{\lambda}_{500}^*$	0.7723	0.4099	0.1657	0.6599	0.2846	0.4176	0.4261	0.2196	0.3984	0.1728

$\hat{\lambda}_N$ : observed downfall probability for normal-AR(1)-GARCH(1,1) model

$\hat{\lambda}_{250}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 250$

$\hat{\lambda}_{500}^*$ : observed downfall probability for calibrated normal-AR(1)-GARCH(1,1) model with  $L = 500$

For all models:  $T = 1000$ , and  $P = 2000$  Value-at-Risk forecasts included

$P_{unc}$  ( $P_{ind}$ ,  $P_{cc}$ ): Probability of observing a sample with higher unconditional coverage test

statistic (independence test statistic, conditional coverage test statistic).

\*/\*\*/\*\*: number of corresponding  $p$ -values smaller than 0.10/0.05/0.01

Table 4: Number of corresponding  $p$ -values below the desired level

Sample:	sig. level: 0.10										sig. level: 0.05										sig. level: 0.01									
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
$\hat{\lambda}_N$	6	8	4	3	7	7	6	7	8	5	6	8	4	3	5	5	6	7	7	5	5	6	2	1	3	4	5	4	6	0
$\hat{\lambda}_{250}^*$	0	2	0	1	1	0	1	0	0	2	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
$\hat{\lambda}_{500}^*$	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\hat{\lambda}_N$	0	9	1	0	7	5	0	2	3	0	0	9	0	0	4	0	0	2	0	0	0	3	0	0	0	0	2	0	0	0
$\hat{\lambda}_{250}^*$	2	10	2	0	5	4	0	3	7	3	1	10	2	0	2	1	0	3	5	1	0	4	0	0	0	0	0	1	1	0
$\hat{\lambda}_{500}^*$	1	9	2	0	4	3	0	3	4	1	0	9	0	0	1	2	0	2	0	0	0	2	0	0	0	0	0	1	0	0
$\hat{\lambda}_N$	6	10	4	3	7	6	5	7	7	4	5	10	3	1	6	6	5	7	6	2	4	8	2	1	4	4	5	4	6	0
$\hat{\lambda}_{250}^*$	1	10	2	1	3	0	0	3	5	2	0	6	0	1	0	0	0	1	1	2	0	2	0	0	0	0	0	0	0	0
$\hat{\lambda}_{500}^*$	0	9	2	0	0	1	0	1	0	1	0	5	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0

investigated, and a slight preference for the longer window length of  $L = 500$  for the simulated data using the  $t$ -A-PARCH model.

The forecasting results we obtain for the five financial return series and the simulated series are quite promising. By design, the method should improve the VaR forecasts of the normal-GARCH model, and indeed, in an extensive out-of-sample forecasting exercise, the unconditional coverage is much better, while the independence of the VaR violations is unaffected by the bias-correction method. More stringently, our proposed method performs better than the  $t$ -GARCH model for almost all considered cases.

Future work could consider the behavior of the proposed methodology for simulated series based on different DGPs, as discussed at the end of Section 3.3, and also extensions of the methodology to the multivariate case. It is well-known that multivariate GARCH models are notoriously problematic because of the large number of parameters, and attempts to constrain them, such as with constant conditional correlations, are untenable in terms of clear empirical evidence (see, e.g., Audrino, 2006, and the references therein, for discussion).

## References

- Alexander, C., 2001. *Market Models: A Guide to Financial Data Analysis*. Wiley & Sons, Chichester.
- Alexander, C., Lazar, E., 2006. Normal mixture garch(1,1): Applications to exchange rate modelling. *Journal of Applied Econometrics* 21, 307–336.
- Audrino, F., 2006. The impact of general non-parametric volatility functions in multivariate GARCH models. *Computational Statistics & Data Analysis* 50, 3032–3052.
- Bams, D., Lehnert, T., Wolff, C. C. P., 2005. An evaluation framework for alternative var-models. *Journal of International Money and Finance* 24, 944–958.
- Bao, Y., Lee, T.-H., Saltoglu, B., 2003. A test for density forecast comparison with applications to risk management. Tech. rep., University of California, Riverside.
- Barone-Adesi, G., Giannopoulos, K., Vosper, L., 1999. Var without correlations for portfolios of derivative securities. *The Journal of Futures Markets* 19 (5), 583–602.
- Barone-Adesi, G., Giannopoulos, K., Vosper, L., 2002. Backtesting derivative portfolios with filtered historical simulation (fhs). *European Financial Management* 8, 31–58.
- Basle Committee on Banking Supervision, 1995. An internal model-based approach to market risk capital requirements, <http://www.bis.org>.
- Basle Committee on Banking Supervision, 1996. Overview of the amendment to the capital accord to incorporate market risks, <http://www.bis.org>.

- Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Broda, S., Paoletta, M. S., 2006. Saddlepoint approximations for the doubly noncentral  $t$  distribution. NCCR Finrisk Working Paper 304, University of Zurich, Zurich, Switzerland.
- Christoffersen, P. F., 1998. Evaluating Interval Forecasts. *International Economic Review* 39 (4), 841–861.
- Christoffersen, P. F., 2003. *Elements of Financial Risk Management*. Academic Press, London.
- Christoffersen, P. F., Goncalves, S., 2005. Estimation Risk in Financial Risk Management. *Journal of Risk* 7, 1–28.
- Ding, Z., Granger, C. W., Engle, R. F., 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Dowd, K., 2005. *Measuring Market Risk*, 2nd Edition. Wiley, New York.
- Dowd, K., Blake, D., 2006. After var: The theory, estimation and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance* 73 (2), 193–229.
- Engel, J., Gizycki, M., 1999. Conservatism, Accuracy and Efficiency: Comparing Value-at-Risk Models. working paper 2, Australian Prudential Regulation Authority.
- Engle, R. F., July 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50 (4), 987–1007.
- Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Frachot, A., 1995. Factor Models of Domestic and Foreign Interest Rates with Stochastic Volatilities. *Mathematical Finance* 5, 167–185.
- Franses, P. H., van Dijk, D., 1996. Forecasting Stock Market Volatility Using (Non-Linear) GARCH Models. *Journal of Forecasting* 15, 229–235.
- Giamouridis, D., 2006. Estimation Risk in Financial Risk Management: A Correction. Forthcoming in: *Journal of Risk*.
- Giot, P., Laurent, S., 2004. Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance* 11, 379–398.
- Gourieroux, C., 1997. *ARCH Models and Financial Applications*. Springer Series in Statistics. Springer-Verlag, New York.
- Gray, S. F., 1996. Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42, 27–62.

- Haas, M., Mittnik, S., Paoletta, M. S., 2004a. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics* 2 (2), 211–250.
- Haas, M., Mittnik, S., Paoletta, M. S., 2004b. A new approach to markov switching garch models. *Journal of Financial Econometrics* 2 (4), 493–530.
- Haas, M., Mittnik, S., Paoletta, M. S., 2006. Modeling and Predicting Market Risk With Laplace-Gaussian Mixture Distributions. To appear in: *Applied Financial Economics*.
- Harvey, C. R., Siddique, A., December 1999. Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis* 34 (4), 465–487.
- He, C., Teräsvirta, T., Malmsten, H., 2002. Fourth Moment Structure of a Family of First-Order Exponential GARCH Models. *Econometric Theory* 18, 868–885.
- Karanasos, M., Kim, J., 2003. Moments of the ARMA-EGARCH Model. *Econometrics Journal* 6, 146–166.
- Klaassen, F., 2002. Improving GARCH Volatility Forecasts with Regime-Switching GARCH. *Empirical Economics* 27, 363–394.
- Kuester, K., Mittnik, S., Paoletta, M. S., 2005. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4 (1), 53–89.
- McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7 (3/4), 271–300.
- Mittnik, S., Paoletta, M. S., 2000. Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates. *Journal of Forecasting* 19, 313–333.
- Mittnik, S., Paoletta, M. S., 2003. Prediction of financial downside-risk with heavy-tailed conditional distributions. In: Rachev, S. T. (Ed.), *Handbook of Heavy Tailed Distributions in Finance*. North-Holland.
- Mittnik, S., Paoletta, M. S., Rachev, S. T., 2000. Diagnosing and Treating the Fat Tails in Financial Returns Data. *Journal of Empirical Finance* 7, 389–416.
- Mittnik, S., Paoletta, M. S., Rachev, S. T., 2002. Stationarity of Stable Power-GARCH Processes. *Journal of Econometrics* 106, 97–107.
- Nelson, D., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59, 347–370.
- Paoletta, M. S., 2001. Testing the Stable Paretian Assumption. *Mathematical and Computer Modelling* 34, 1095–1112.
- Pascual, L., Romo, J., Ruiz, E., 2006. Bootstrap prediction for returns and volatilities in garch models. *Computational Statistics & Data Analysis* 50, 2293–2312.

Rachev, S. T., Mittnik, S., 2000. *Stable Paretian Models in Finance*. Wiley & Sons, Chichester.

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

## CFS Working Paper Series:

No.	Author(s)	Title
2006/23	<b>Christoph Hartz</b> <b>Stefan Mittnik</b> <b>Marc S. Paoella</b>	Accurate Value-at-Risk Forecasting Based on the (good old) Normal-GARCH Model
2006/22	<b>Dirk Krueger</b> <b>Hanno Lustig</b> <b>Fabrizio Perri</b>	Evaluating Asset Pricing Models with Limited Commitment using Household Consumption Data
2006/21	<b>Juan Carlos Conesa</b> <b>Sagiri Kitao</b> <b>Dirk Krueger</b>	Taxing Capital? Not a Bad Idea After All!
2006/20	<b>Annamaria Lusardi</b> <b>Olivia S. Mitchell</b>	Baby Boomer Retirement Security: The Roles of Planning, Financial Literacy, and Housing Wealth
2006/19	<b>Carol C. Bertaut</b> <b>Michael Haliassos</b>	Credit Cards: Facts and Theories
2006/18	<b>Dirk Krueger</b> <b>Alexander Ludwig</b>	On the Consequences of Demographic Change for Rates of Returns to Capital, and the Distribution of Wealth and Welfare
2006/17	<b>Franklin Allen</b> <b>Elena Carletti</b>	Mark-to-Market Accounting and Liquidity Pricing
2006/16	<b>Erik Hurst</b> <b>Arthur Kennickell</b> <b>Annamaria Lusardi</b> <b>Francisco Torralba</b>	Precautionary Savings and the Importance of Business Owners
2006/15	<b>Arthur Kennickell</b> <b>Annamaria Lusardi</b>	Disentangling the Importance of the Precautionary Saving Motive
2006/14	<b>Yanis Biliias,</b> <b>Dimitris Georgarakos</b> <b>Michael Haliassos</b>	Portfolio Inertia and Stock Market Fluctuations

Copies of working papers can be downloaded at <http://www.ifk-cfs.de>