



HAL
open science

GARCH-UGH: a bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series

Hibiki Kaibuchi, Yoshinori Kawasaki, Gilles Stupfler

► To cite this version:

Hibiki Kaibuchi, Yoshinori Kawasaki, Gilles Stupfler. GARCH-UGH: a bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series. *Quantitative Finance*, 2022. hal-04061113

HAL Id: hal-04061113

<https://hal.science/hal-04061113>

Submitted on 6 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–15

GARCH-UGH: a bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series

H. KAIBUCHI * †, Y. KAWASAKI † and G. STUPFLER ‡

† The Graduate University for Advanced Studies, The Institute of Statistical Mathematics - Tachikawa, Tokyo, Japan

‡ Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France

(Received 00 Month 20XX; in final form 00 Month 20XX)

The Value-at-Risk (VaR) is a widely used instrument in financial risk management. The question of estimating the VaR of loss return distributions at extreme levels is an important question in financial applications, both from operational and regulatory perspectives; in particular, the dynamic estimation of extreme VaR given the recent past has received substantial attention. We propose here a new two-step bias-reduced estimation methodology for the estimation of one-step ahead dynamic extreme VaR, called GARCH-UGH (Unbiased Gomes-de Haan), whereby financial returns are first filtered using an AR-GARCH model, and then a bias-reduced estimator of extreme quantiles is applied to the standardized residuals. Our results indicate that the GARCH-UGH estimates of the dynamic extreme VaR are more accurate than those obtained either by historical simulation, conventional AR-GARCH filtering with Gaussian or Student- t innovations, or AR-GARCH filtering with standard extreme value estimates, both from the perspective of in-sample and out-of-sample backtestings of historical daily returns on several financial time series.

Keywords: Bias correction · Extreme value theory (EVT) · Financial time series · GARCH model · Hill estimator · Value-at-Risk (VaR)

JEL Classification: C1, C32, C53, G1

1. Introduction

A major concern in financial risk management is to quantify the risk associated to high-impact, low-probability extreme losses. The most widely known risk measure is Value-at-Risk (VaR), defined as a quantile of the loss distribution. Even though the Basel Committee on Banking Supervision recommends the use of VaR at high levels (see for example Basel Committee on Banking Supervision 2013), it has been criticized several times in the financial literature for two main reasons. First, the VaR only measures the frequency of observations below or above the predictor and not their magnitude: this means that, while it is known that $100(1 - \tau)\%$ of losses will be higher than the VaR q_τ at level τ , the VaR alone cannot give any further information about the size of these large losses. Second, the VaR is not a coherent risk measure in the sense of Artzner *et al.* (1999), because it is not sub-additive in general, meaning that it does not abide by the intuitive diversification principle stating that a portfolio built on several financial assets carries less risk than a portfolio solely consisting of one of these assets. These two weaknesses pushed the Basel Committee to also

*Corresponding author. Email: kaibuchi@ism.ac.jp

recommend calculating the Expected Shortfall (or Conditional Value-at-Risk) as a complement or alternative to the VaR. In practice, this is hampered by the fact that the Expected Shortfall is not elicitable [in the sense of Gneiting \(2011\)](#), and therefore the development of a simple backtesting methodology for the Expected Shortfall is not clear [\(see Deng and Qiu 2021 for a very recent comprehensive study of backtesting procedures for the Expected Shortfall\)](#).

This is why the accurate estimation of VaR is worth pursuing, and is our focus in this paper. An estimation of the unconditional VaR (that is, of the common distribution of returns over time, assumed to be stationary) is appropriate for the estimation of potential large levels of loss over the long term, for example with the goal of making long-term investment decisions. On the other hand, the conditional VaR is more appropriate for day-to-day and short-term risk management by capturing the dynamics and the key properties of financial asset returns such as volatility clustering and leptokurtosis. The estimation of the extreme conditional VaR, on which we focus in this paper, therefore gives a better understanding of the riskiness of the portfolio because this riskiness varies with the changing volatility. Quantifying the risk associated to extreme losses can then be done by estimating an extreme quantile of order $1 - p$, where $p = p(n) \rightarrow 0$ as the available sample size n of the data tends to infinity.

There are two main classes of methods to estimate the conditional VaR. Nonparametric historical simulation (HS) relies on observed data directly, and uses the empirical distribution of past losses without assuming any specific distribution, see Danielsson (2011) and McNeil *et al.* (2015). Although HS is easy to implement, the estimation of extreme quantiles using HS is difficult as the extrapolation beyond observed returns is impossible because this method essentially assumes that one of the observed returns is expected to be the next period return. By contrast, the parametric approach generally refers to the use of an econometric model of volatility dynamics such as, among many others, the Generalized Autoregressive Conditional heteroskedasticity (GARCH) model of Bollerslev (1986). These models estimate VaRs reflecting the conditional heteroskedasticity of financial data. However, GARCH-type models assuming a parametric distribution of the innovation variable tend to underestimate risk at extreme levels because this strong parametric assumption is not well-suited to the accurate estimation of heavy-tailedness in the conditional returns of financial time series.

To overcome the problems of purely nonparametric or parametric estimations of extreme VaR, McNeil and Frey (2000) proposes a two-step approach combining a GARCH-type model and Extreme Value Theory (EVT), referred to as GARCH-EVT throughout. EVT focuses only on the tails and allows the extrapolation beyond available data, while traditional econometric models focus on the whole distribution at the expense of consideration of the tails. The key idea of the GARCH-EVT method is to estimate the dynamic extreme VaR by first filtering financial time series with a GARCH-type model to estimate the current volatility, and then by applying the EVT method to the standardized residuals for estimating the tails of the residual distribution. This approach has been widely used to estimate the extreme conditional VaR. For instance Byström (2004) and Fernandez (2005) find GARCH-EVT to give accurate VaR estimates for standard and extreme quantiles compared with GARCH-type models and unconditional EVT methods on stock market data collected across the US, Latin America, Europe and Asia. Being a two-step procedure based on GARCH-type filtering, the accuracy of the GARCH-EVT approach has been debated. Chavez-Demoulin *et al.* (2005) points out that estimates of the extreme conditional VaR via the GARCH-EVT approach are sensitive to the fitting of a GARCH-type model to the dataset in the first step. On the other hand, Furió and Climent (2013) and Jalal and Rockinger (2008) have concluded that there is no evidence of any difference in the final VaR estimates, regardless of the particular GARCH model selected to filter financial data.

Since the debate on filtering, several modifications of the conventional GARCH-EVT method have been suggested in the literature to provide a more accurate calculation of the residuals before applying the EVT method in the second step. Yi *et al.* (2014) proposes a semiparametric version of GARCH-EVT based on quantile regression. Youssef *et al.* (2015) adapts the FIGARCH, HY-GARCH and FIAPARCH models to estimate extreme conditional VaRs for crude oil and gasoline

market. Bee *et al.* (2016) proposes an approach called realized EVT where returns are pre-whitened with a high-frequency based volatility model. Zhao *et al.* (2019) develops hybrid time-varying long-memory GARCH-EVT models by using a variety of fractional GARCH models. To the best of our knowledge, little work has been carried out on the EVT step itself; Ergen (2015) uses the skewed t -distribution that is fitted to the standardized residuals from the GARCH step in order to recover a fully parametric specification. In the context of estimation and inference of unconditional extreme VaR, bias correction is a key concern and has an extensive history (see *e.g.* Cai *et al.* 2013 for a review). de Haan *et al.* (2016) develop a semiparametric bias-reduced estimator of extreme *unconditional* VaR in stationary time series. However, such improvements have not been investigated so far in the specific context of *dynamic* estimation of extreme quantiles of financial time series.

This is the contribution of the present paper. More precisely, in the context of the estimation of the one-step ahead dynamic extreme VaR, we develop a new methodology called GARCH-UGH (standing for Unbiased Gomes-de Haan, after de Haan *et al.* 2016 and Gomes *et al.* 2002). The novelty in this methodology is that, instead of applying the Peaks-Over-Threshold (POT) method in the GARCH-EVT approach as in McNeil and Frey (2000), we use an asymptotically unbiased estimator, derived from the work of de Haan *et al.* (2016), of the extreme quantile applied to the standardized residuals from the GARCH step. We analyze the performance of our approach on four financial time series, which are the Dow Jones, Nasdaq and Nikkei stock indices, and the Japanese Yen/British Pound exchange rate. As we shall illustrate, our results indicate that GARCH-UGH provides substantially more accurate one-step ahead extreme conditional VaRs than either HS, the conventional GARCH-N method (that is, the standard GARCH specification of heteroskedasticity with normal innovations), its GARCH- t analogue where the innovations are Student- t distributed, the GARCH-EVT approach, or bias-reduced EVT without filtering, based on the performance of the in-sample and out-of-sample backtestings. In addition, our bias-reduction procedure will be designed to be robust to departure from the independence assumption, and as such will be able to handle residual dependence present after filtering in the first step. Our finite-sample results will also illustrate that the GARCH-UGH method leads to one-step ahead extreme conditional VaR estimates that are less sensitive to the choice of sample fraction, and hence mitigates the difficulty in selecting the optimal number of observations for the estimations. Finally, the computational cost of GARCH-UGH is lower than that of conventional GARCH-EVT: the extreme value step in the GARCH-UGH method is semiparametric with an automatic and fast recipe for the estimations of the one-step ahead extreme conditional VaRs, while the GARCH-EVT method is based on a parametric fit of the Generalized Pareto Distribution (GPD) to the residuals using Maximum Likelihood Estimation.

The rest of the paper is organized as follows. Section 2 presents our framework and methodology. Section 3 first describes the four financial time series used in the empirical analysis, and discusses the performance of our proposed approach through in-sample and out-of-sample backtestings of one-step ahead extreme VaRs compared to existing approaches. Section 4 discusses our findings and perspectives for future work.

2. The GARCH-UGH method and framework

Let p_t be a daily-recorded price for a stock, index or exchange rate, and let $X_t = -\log(p_t/p_{t-1})$ be the negative daily log-return on this price. We assume that the dynamics of X_t are governed by

$$X_t = \mu_t + \sigma_t Z_t, \quad (1)$$

where $\mu_t \in \mathbb{R}$ and $\sigma_t > 0$ denote the (conditional) mean and standard deviation, and the innovations Z_t form a strictly stationary white noise process, that is, they are i.i.d. with zero mean, unit variance and common marginal distribution function F_Z . We assume that for each t , μ_t and σ_t are measurable with respect to the σ -algebra \mathcal{F}_{t-1} representing the information about the return

process available up to time $t - 1$.

We are concerned with estimating extreme conditional quantiles of these negative log-returns. Recall that for a probability level $\tau \in (0, 1)$, the τ th unconditional quantile of a distribution F is $q_\tau = \inf\{x \in \mathbb{R} : F(x) \geq \tau\}$. Here we focus on the one-step ahead quantile, that is, the estimation of the conditional extreme quantile of X_{t+1} given \mathcal{F}_t , whose order τ tends to 1 as the available sample size n goes to infinity. In this case, by location equivariance and positive homogeneity of quantiles, the one-step ahead conditional quantile (or VaR) of X_{t+1} can be written as

$$q_\tau(X_{t+1} \mid \mathcal{F}_t) = \mu_{t+1} + \sigma_{t+1}q_\tau(Z), \quad (2)$$

where $q_\tau(Z)$ is the common τ th quantile of the marginal distribution of the innovations Z_t . The problem of estimating $q_\tau(X_{t+1} \mid \mathcal{F}_t)$ can then be tackled by estimating the mean and standard deviation components μ_{t+1} and σ_{t+1} and the unconditional quantile $q_\tau(Z)$. Given estimates $\hat{\mu}_{t+1}$, $\hat{\sigma}_{t+1}$ and $\hat{q}_\tau(Z)$ of these quantities, an estimate of $q_\tau(X_{t+1} \mid \mathcal{F}_t)$ is then

$$\hat{q}_\tau(X_{t+1} \mid \mathcal{F}_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}_\tau(Z).$$

In calculating this estimate, there are three main difficulties. First, one has to estimate μ_{t+1} and σ_{t+1} , which supposes that an appropriate model and estimation method have to be chosen. Second, the innovations Z_t are unobserved, which means that the estimation of $q_\tau(Z)$ has to be based on residuals following the estimation of μ_{t+1} and σ_{t+1} . A third difficulty is specific to our context: we wish here to estimate a dynamic extreme VaR, that is, a conditional quantile $q_\tau(X_{t+1} \mid \mathcal{F}_t)$ with τ very close to 1. In such contexts, it is well-known that traditional nonparametric estimators become inconsistent (see for example the monographs by Beirlant *et al.* 2004 and Embrechts *et al.* 1997), and adapted extrapolation methodologies have to be employed.

Our GARCH-UGH method combines estimation of the mean and standard deviation in a GARCH-type model with a flexible bias-reduced extrapolation methodology for the estimation of $q_\tau(Z)$ ($\tau \uparrow 1$) using the residuals obtained after estimation of the model structure. We describe these two steps successively below.

2.1. GARCH step

In order to estimate μ_{t+1} and σ_{t+1} , one should select a particular model in the class (1). Many different models for volatility dynamics have been used in the literature of GARCH-EVT approach, as we highlighted in our literature review in Section 1. Here we use an AR(1) model for the dynamics of the conditional mean, and a parsimonious but effective GARCH(1,1) model for the volatility, as in the original GARCH-EVT approach; this will allow us to subsequently illustrate how improving the second, EVT-based step can result in more accurate estimates of extreme conditional VaR.

We thus model the conditional mean of the series by $\mu_t = \phi X_{t-1}$, for some $\phi \in (-1, 1)$, and the conditional variance of the mean-adjusted series $\epsilon_t = X_t - \mu_t$ by $\sigma_t^2 = \kappa_0 + \kappa_1 \epsilon_{t-1}^2 + \kappa_2 \sigma_{t-1}^2$, where $\kappa_0, \kappa_1, \kappa_2 > 0$. Necessary and sufficient conditions for the stationarity of a model following GARCH(1,1) dynamics are given in Chapter 2 of Francq and Zakoïan (2010); the condition $\kappa_1 + \kappa_2 < 1$ is a simple sufficient condition guaranteeing stationarity. The model is therefore the AR(1)-GARCH(1,1) model

$$X_t = \mu_t + \sigma_t Z_t, \quad \text{with } \mu_t = \phi X_{t-1} \text{ and } \sigma_t^2 = \kappa_0 + \kappa_1 (X_{t-1} - \mu_{t-1})^2 + \kappa_2 \sigma_{t-1}^2. \quad (3)$$

In Equation (3), the innovations Z_t are i.i.d. with zero mean, unit variance.

In order to make one-step ahead predictions at time t , we fix a memory n so that at the end of time t , the financial data consist of the last n negative log-returns X_{t-j} , for $0 \leq j \leq n - 1$. We then fit the AR(1)-GARCH(1,1) model to the data $(X_{t-n+1}, \dots, X_{t-1}, X_t)$ using Gaussian Quasi-Maximum Likelihood Estimation (QMLE), that is, by maximizing the likelihood constructed by

assuming that the innovations Z_t are i.i.d. Gaussian with zero mean and unit variance. **The R package rugarch (Galanos 2019) has been used for the estimation.** While of course the innovations Z_t will not be Gaussian in general (and indeed in our UGH step we shall assume that they are heavy-tailed), the QMLE method yields a consistent and asymptotically normal estimator, see for example Francq and Zakoian (2004) for a theoretical analysis. One may also put a strong heavy-tailed parametric specification on Z_t , such as assuming that they are location-scale Student distributed; this was tried in our analysis of financial log-returns but did not improve results substantially.

Let $(\hat{\phi}, \hat{\kappa}_0, \hat{\kappa}_1, \hat{\kappa}_2)$ be the Gaussian QMLE estimates. Choosing sensible starting values for $\hat{\epsilon}_{t-n}^2$ and $\hat{\sigma}_{t-n}^2$ (for example, constant values as in Section 7.1 of Francq and Zakoian (2010)), estimates of the conditional mean and the conditional standard deviation, $(\hat{\mu}_{t-n+1}, \dots, \hat{\mu}_{t-1}, \hat{\mu}_t)$ and $(\hat{\sigma}_{t-n+1}, \dots, \hat{\sigma}_{t-1}, \hat{\sigma}_t)$ respectively, can be calculated from Equation (3) recursively. This leads to the residuals

$$(\hat{Z}_{t-n+1}, \dots, \hat{Z}_t) = \left(\frac{X_{t-n+1} - \hat{\mu}_{t-n+1}}{\hat{\sigma}_{t-n+1}}, \dots, \frac{X_t - \hat{\mu}_t}{\hat{\sigma}_t} \right).$$

We end this step by calculating the estimates of the conditional mean and standard deviation for time $t + 1$, which are the obvious one-step ahead forecasts, as follows:

$$\begin{aligned} \hat{\mu}_{t+1} &= \hat{\phi} X_t, \\ \hat{\sigma}_{t+1} &= \sqrt{\hat{\kappa}_0 + \hat{\kappa}_1 \hat{\epsilon}_t^2 + \hat{\kappa}_2 \hat{\sigma}_t^2}, \end{aligned}$$

where $\hat{\epsilon}_t = X_t - \hat{\mu}_t$. In summary, this first GARCH step of the method consists in fitting an AR(1)-GARCH(1,1) model to the negative log-returns at a certain past time horizon n (not too small so that the method produces reasonable results, and not too large so that the AR-GARCH model is believable over this time period), using a Gaussian QMLE, leading to forecasts $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ and standardized residuals \hat{Z}_{t-j} , $0 \leq j \leq n - 1$.

2.2. UGH step

With standardized residuals at our disposal, we can now discuss the estimation of the extreme quantile $q_\tau(Z)$ of the innovations Z_t , for $\tau \uparrow 1$. The residuals \hat{Z}_{t-j} , $0 \leq j \leq n - 1$, approximate the true unobservable Z_{t-j} . Assume that the underlying distribution of these Z_{t-j} is heavy-tailed, that is (see Theorem 1.2.1 p.19 and Corollary 1.2.10 p.23 in de Haan and Ferreira 2006)

$$\lim_{t \rightarrow \infty} \frac{U(tz)}{U(t)} = z^\gamma, \quad \forall z > 0, \quad \text{where } U(t) = q_{1-t^{-1}}(Z). \quad (4)$$

In other words, we assume the tail of the innovations to be approximately Pareto, with the so-called extreme value index γ tuning how heavy the tail is. This assumption is ubiquitous in actuarial and financial risk management (see e.g. p.9 of Embrechts *et al.* 1997 and p.1 of Resnick 2007). It makes it possible to construct extrapolated extreme quantile estimators: the classical Weissman quantile estimator (see Weissman 1978) of a quantile $q_\tau(Z) = q_{1-p}(Z)$ with $p = 1 - \tau$ close to 0 (meaning, in mathematical terms, that $p = p(n) \rightarrow 0$ as $n \rightarrow \infty$) is then

$$\bar{q}_{1-p}(Z) = \left(\frac{k}{np} \right)^{\bar{\gamma}_k} Z_{n-k,n} \quad (5)$$

where $Z_{1,n} \leq Z_{2,n} \leq \dots \leq Z_{n,n}$ are the order statistics from Z_{t-n+1}, \dots, Z_t and $\bar{\gamma}_k$ is a consistent estimator of γ . The tuning parameter k denotes the effective sample size for the estimation: this

parameter should be chosen not too small, so that the variance of the estimator is reasonable, but also not too large so that the bias coming from the use of the extrapolation relationship (4) does not dominate. The most common estimator $\bar{\gamma}_k$ of γ is the Hill estimator (introduced in Hill 1975):

$$\bar{\gamma}_k = \bar{\gamma}_k^H = \frac{1}{k} \sum_{i=1}^k \log Z_{n-i+1,n} - \log Z_{n-k,n}. \quad (6)$$

The Hill and Weissman estimators can be shown to be asymptotically Gaussian under suitable conditions on $k = k(n)$ (see for example Chapters 3 and 4 in de Haan and Ferreira 2006). A reasonable idea to define an estimator of $q_{1-p}(Z)$ in our context is then to use the estimators defined in Equations (5) and (6) with the order statistics of the residuals, $\hat{Z}_{n-j,n}$, in place of the unobservable $Z_{n-j,n}$.

The choice of the parameter k requires solving a bias-variance tradeoff for which there is no straightforward approach. Indeed, with a low k , the estimators use observations that are very informative about the extremes, but their low number results in a high variance. With a high k , the variance is reduced, but at the cost of taking into account observations that are further into the bulk of the distribution and thus carry bias. One possible way to make the choice of k easier is to work on correcting this bias. This can be done under the following so-called second-order condition on U :

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \left(\frac{U(tz)}{U(t)} - z^\gamma \right) = z^\gamma \frac{z^\rho - 1}{\rho}, \quad \forall z > 0, \quad (7)$$

where $\rho \leq 0$ is called the second-order parameter and A is a positive or negative function converging to 0 at infinity, such that $|A|$ is regularly varying with index ρ . See Equation (6.35) p.341 in Embrechts *et al.* (1997) and, in our parametrization, Theorem 2.3.9 p.48 in de Haan and Ferreira (2006). The function A therefore controls the rate of convergence in Equation (4): the larger $|\rho|$ is, the faster $|A|$ converges to 0, and the smaller the error in the approximation of the right tail of U by a Pareto tail is. This makes it possible to precisely quantify the bias of the Hill and Weissman estimators, and to correct for this bias by estimating the function A and the parameter ρ . This results in bias-corrected Hill and Weissman estimators for which the selection of k is typically much easier because their performance is much more stable.

Our idea in this second, UGH step is to apply such bias-corrected estimators constructed in de Haan *et al.* (2016) (and built on second-order parameter estimators of Gomes *et al.* 2002, hence the name UGH, for Unbiased Gomes-de Haan) to our residuals obtained from the GARCH step. Our estimator of ρ motivated by Gomes *et al.* (2002) is

$$\hat{\rho}_k^{(\alpha)} = (s^{(\alpha)})^{\leftarrow} (S_k^{(\alpha)}).$$

Here $\alpha \notin \{1/2, 1\}$ is a positive tuning parameter, $s^{(\alpha)\leftarrow}$ denotes the generalized (left-continuous) inverse of the function

$$s^{(\alpha)}(\rho) = \frac{\rho^2(1 - (1 - \rho)^{2\alpha} - 2\alpha\rho(1 - \rho)^{2\alpha-1})}{(1 - (1 - \rho)^{\alpha+1} - (\alpha + 1)\rho(1 - \rho)^\alpha)^2},$$

and we set

$$S_k^{(\alpha)} = \frac{\alpha(\alpha+1)^2 \Gamma^2(\alpha)}{4 \Gamma(2\alpha)} \frac{R_k^{(2\alpha)}}{(R_k^{(\alpha+1)})^2},$$

with $R_k^{(\alpha)} = \frac{M_k^{(\alpha)} - \Gamma(\alpha+1)(M_k^{(1)})^\alpha}{M_k^{(2)} - 2(M_k^{(1)})^2}$ and $M_k^{(\alpha)} = \frac{1}{k} \sum_{i=1}^k (\log \hat{Z}_{n-i+1,n} - \log \hat{Z}_{n-k,n})^\alpha$.

The version of $\hat{\rho}_k^{(\alpha)}$ with the true innovations Z_k instead of the residuals is known to be consistent under a so-called third-order condition which further strengthens (7). Here we choose $\alpha = 2$ since, on fully observed data, this appears to yield the smallest mean squared error following the numerical simulations of Gomes *et al.* (2002). This results in the estimator

$$\hat{\rho}_k^{(2)} = \frac{-4 + 6S_k^{(2)} + \sqrt{3S_k^{(2)} - 2}}{4S_k^{(2)} - 3}$$

provided $2/3 \leq S_k^{(2)} \leq 3/4$, where

$$S_k^{(2)} = \frac{3}{4} \frac{[M_k^{(4)} - 24(M_k^{(1)})^4][M_k^{(2)} - 2(M_k^{(1)})^2]}{[M_k^{(3)} - 6(M_k^{(1)})^3]^2}.$$

[This expression corrects a typo in p.375 of de Haan *et al.* (2016).] It is clear that $\hat{\rho}_k^{(2)}$ does not exist if $S_k^{(2)} \notin [2/3, 3/4]$. In practice, we select the value of k in this estimator by setting

$$k_\rho = \sup \left\{ k : k \leq \min \left(m - 1, \frac{2m}{\log \log m} \right) \text{ and } \hat{\rho}_k^{(2)} \text{ exists} \right\}. \quad (8)$$

Here m is the number of positive observations in the sample. The intuition is that even though this estimator of ρ requires a choice of k , this choice should be different from its counterpart used in the estimation of γ , and indeed intuitively the value of k in the estimator of ρ should be rather high in order to allow the methodology to identify the bias coming from including observations belonging to the bulk of the distribution (which correspond to a high k). We then estimate ρ by $\hat{\rho}_{k_\rho} = \hat{\rho}_{k_\rho}^{(2)}$. If the set on the right-hand side of (8) is empty, we define $\hat{\rho}_{k_\rho} = -1$ as recommended in p.117 of Section 4.5.1 in Beirlant *et al.* (2004), in p.212-215 in Gomes *et al.* (2000) and in p.195 of Section 6.6 in Reiss and Thomas (2007). The choice $\hat{\rho}_{k_\rho} = -1$ is an *ad hoc* compromise between bias reduction and variability of the estimators; note that the estimation of the constant ρ is known to be a difficult problem in finite samples (see Gomes *et al.* 2009, p.298 and Goegebeur *et al.* 2010, p.2638, where it is seen that estimators of ρ typically have a low rate of convergence). The parameter γ is then estimated using the residual-based version of the bias-corrected Hill estimator introduced in de Haan *et al.* (2016):

$$\hat{\gamma}_{k,k_\rho} = \hat{\gamma}_k^H - \frac{M_k^{(2)} - 2(\hat{\gamma}_k^H)^2}{2\hat{\gamma}_k^H \hat{\rho}_{k_\rho} (1 - \hat{\rho}_{k_\rho})^{-1}}.$$

The addend in the right-hand side corresponds to the bias correction. We now have all the necessary ingredients to define our residual-based, bias-corrected estimator of unconditional extreme quantiles

of Z :

$$\hat{q}_{1-p}(Z) = \left(\frac{k}{np}\right)^{\hat{\gamma}_{k,k_\rho}} \hat{Z}_{n-k,n} \times \left(1 - \frac{[M_k^{(2)} - 2(\hat{\gamma}_k^H)^2][1 - \hat{\rho}_{k_\rho}]^2}{2\hat{\gamma}_k^H \hat{\rho}_{k_\rho}^2} \left[1 - \left(\frac{k}{np}\right)^{\hat{\rho}_{k_\rho}}\right]\right). \quad (9)$$

This corresponds to a slightly different version of the estimator in Section 4.3 of de Haan *et al.* (2016), given later by Chavez-Demoulin and Guillou (2018), who pointed out a mistake in the analysis of de Haan *et al.* (2016). The use of $\hat{\gamma}_{k,k_\rho}$ rather than the Hill estimator in the extrapolation will correct the bias due to the estimation of the extreme value index; the multiplier corrects the bias specifically due to the use of the Pareto distribution for the extrapolation of extreme quantiles. The versions of these estimators for fully observed data work when this data is weakly serially dependent, as shown in Chavez-Demoulin and Guillou (2018). As such, our proposed method will be robust to the presence of residual dependence after filtering and to model misspecification in the sense of Hill (2015). We shall also show that the choice of k for this estimator is not as crucial in finite samples as for the traditional Hill estimator, because this estimator has reasonably good performance across a large range of values of k .

Note that if the Z_{t-j} , $0 \leq j \leq n-1$ are independent, then Theorem 4.2 of de Haan *et al.* (2016) suggests that

$$\frac{\sqrt{k}}{\log(k/np)} \left(\frac{\hat{q}_{1-p}(Z)}{q_{1-p}(Z)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\gamma^2}{\rho^2} (\rho^2 + (1-\rho)^2)\right).$$

A standard Gaussian 95% asymptotic confidence interval for the extreme quantile $q_{1-p}(Z)$, $p \downarrow 0$ is then given by

$$\left[\hat{q}_{1-p}(Z) \left(1 \pm \frac{1.96}{\sqrt{k}/\log(k/np)} \times \sqrt{\frac{\hat{\gamma}_{k,k_\rho}^2}{\hat{\rho}_{k_\rho}^2} (\hat{\rho}_{k_\rho}^2 + (1 - \hat{\rho}_{k_\rho})^2)} \right) \right].$$

This asymptotic Gaussian confidence interval is easy to implement, but of course its validity relies on assuming that the negative log-returns are correctly filtered using the AR-GARCH model.

2.3. Summary and output of the GARCH-UGH method

The GARCH-UGH approach may be briefly summarized by the following two successive steps:

- (i) GARCH step: based on n previous observations at time t , fit an AR(1)-GARCH(1,1) model to the negative daily log-returns data using a Gaussian QMLE. Obtain $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ using the fitted model and compute standardized residuals. See Section 2.1 for full details on this GARCH step, similar to the first step in McNeil and Frey (2000).
- (ii) UGH step: use these standardized residuals as proxies for the true unobserved innovations Z_{t-j} , $0 \leq j \leq n-1$, to construct the asymptotically unbiased tail quantile estimator $\hat{q}_{1-p}(Z)$ in (9). See Section 2.2 for full details on this UGH step which is a different, residual-based new version of the procedure of de Haan *et al.* (2016).

Combining the two steps results in the final GARCH-UGH estimator

$$\hat{q}_\tau(X_{t+1} | \mathcal{F}_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{q}_\tau(Z), \quad \tau = 1-p, \quad p \text{ close to } 0.$$

The goal of our real data analysis is to examine the finite-sample performance of this estimator for low exceedance probabilities, that is, τ close to 1.

3. Empirical analysis of four financial time series

We consider historical daily negative log-returns of three financial indices and an exchange rate, all made of $n = 4000$ observations:

- The Dow Jones Industrial Average (DJ) from 23 December 1993 to 9 November 2009;
- The Nasdaq Stock Market Index (NASDAQ) from 30 August 1993 to 16 July 2009;
- The Nikkei 225 (NIKKEI) from 14 May 1993 to 12 August 2009;
- The Japanese Yen-British Pound exchange rate (JPY/GBP) from 2 January 2000 to 14 December 2010.

The data have been taken from the R package `qrmdata` (Hofert and Hornik 2016) and are represented in Figure 1. The graphs show that these negative log-returns are extremely volatile around the 2007-2008 financial crisis, which created a succession of extreme positive and negative returns over short time horizons. A noticeable degree of volatility clustering is also detected from a visual inspection of Figure 1, revealing the presence of heteroskedasticity. Including a turbulent period from a financial risk management perspective is crucial in order to examine how dynamic extreme VaR estimators behave. An inspection of more recent financial data collected during the COVID-19 crisis did not reveal a more substantial degree of volatility, so we focus on the well-studied subprime crisis in order to assess the quality of our forecasts.

Descriptive statistics and basic statistical tests applied to the negative log-returns on the four financial time series are reported in Table 1. According to the descriptive statistics, the mean of the negative log-returns of all series are close to zero, and negative log-returns are leptokurtic. The Jarque-Bera test statistics indicate that the Gaussian distribution is not suitable for any of these series of negative log-returns. All four series pass the augmented Dickey-Fuller (ADF) test, indicating that they can be considered stationary for modeling purposes. The Ljung-Box test applied to the squared negative log-returns, with orders 1 and 10, rejects the null hypothesis of no autocorrelation, indicating the presence of substantial conditional heteroskedasticity in all series. This provides justification for our use of GARCH-type models with these data.

We compare six methods in total:

- The nonparametric HS method is based on the observed data, and its VaR is simply the empirical quantile of the series X_t at the desired quantile level.
- The GARCH-N (normal) method uses the same filtering step as explained in Section 2.1, but assumes in the quantile estimation step also that the innovations Z_t are i.i.d. $\mathcal{N}(0, 1)$. The extreme conditional VaR is then calculated as $\hat{q}_\tau(X_{t+1} | \mathcal{F}_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \Phi^{-1}(\tau)$, where Φ is the cumulative distribution function of the standard normal distribution.
- The GARCH- t method again uses the filtering step of Section 2.1, but assumes in the quantile estimation step that the standardized residuals from the GARCH step are Student- t distributed. This corresponds to the two-step estimation method discussed on p.1017 of Ergen (2015).
- The bias-reduced UGH method without filtering: this method applies the UGH step directly to the series X_t .
- The conventional GARCH-EVT method as described in McNeil and Frey (2000). This consists, first, in the same filtering step as described in Section 2.1. Standardized residuals are then recorded and a Generalized Pareto distribution is fitted using a maximum likelihood estimator, thus producing a VaR estimate $\hat{q}_\tau(Z)$. This method therefore differs from ours as far as the extreme value step is concerned.
- The proposed GARCH-UGH method.

A comparison with the basic estimation methods (HS, GARCH-N and GARCH- t) indicates the importance of extreme value methods in the estimation of the dynamic extreme VaR. Besides, a comparison with the UGH method (without filtering) allows us to see how effective filtering is, and a comparison with the GARCH-EVT method (not featuring bias reduction) will illustrate the

benefit of bias reduction at the extreme value step after filtering. In Section 3.1, we explain how we carry out backtesting of the performance of one-step ahead extreme conditional VaR estimators provided by each approach. This is then followed by in-sample and out-of-sample evaluations of one-step ahead conditional VaR estimates at different τ levels and choices of k in Sections 3.2 and 3.3, respectively: in-sample estimation investigates the fit of the approaches to high volatile returns, while out-of-sample estimation tests how well the method predicts extreme VaR.

3.1. Statistical framework for VaR backtesting

Backtesting is carried out to examine the accuracy of the one-step ahead conditional VaR estimates. It compares the ex-ante VaR estimates $\hat{q}_\tau(X_t | \mathcal{F}_{t-1})$ with the ex-post realized negative log-returns in a time window W_T , with a VaR violation at time t said to occur whenever $x_t > \hat{q}_\tau(X_t | \mathcal{F}_{t-1})$. Define a hit sequence of VaR violations as $I_t = \mathbb{1}\{x_t > \hat{q}_\tau(X_t | \mathcal{F}_{t-1})\}$. If a VaR estimation method is accurate, then the sequence (I_t) should approximately be an independent sequence of Bernoulli variables with success probability $p = 1 - \tau$. Both the distributional and independence properties are equally important. A VaR estimation method with too few VaR violations will tend to overestimate risk and therefore to be excessively conservative in financial terms, while too many VaR violations mean that risk is underestimated, leading to insufficient provision of capital and therefore potential insolvency in case of large losses. Besides, a violation of the independence property typically arises when there is a clustering of VaR violations, which indicates a model that does not represent volatility clustering well enough.

In order to test the distributional assumption, we use the unconditional likelihood ratio coverage test proposed by Kupiec (1995), also known as Kupiec test or POF test, for Proportion Of Failures. To test the independence property, we use another likelihood ratio test called the conditional coverage test, proposed in Christoffersen (1998) and also known as the Christoffersen test, based on testing for first-order Markov dependence. Strictly speaking the conditional coverage test only assesses departure from either independence or stationarity, but in fact the test statistic is the sum of the unconditional coverage test statistic LR_{uc} and a likelihood ratio test statistic of independence LR_{ind} , so checking stationarity and independence via the pair of test statistics (LR_{uc}, LR_{ind}) is exactly equivalent to checking them via the unconditional and conditional coverage tests. We therefore use below both the unconditional and conditional coverage tests to assess the performance of our dynamic extreme VaR estimators. This constitutes a backtesting approach in the spirit of the one suggested by the Basel Committee on Banking Supervision.

3.2. In-sample dynamic extreme VaR estimation and backtesting

We start by estimating in-sample one-step ahead conditional extreme VaRs $q_\tau(X_{t+1} | \mathcal{F}_t)$ for $\tau \in \{0.99, 0.995, 0.999\}$. For these in-sample evaluations, all methods (HS, GARCH-N and GARCH-t, UGH without filtering, GARCH-EVT without bias reduction, and our proposed GARCH-UGH method) are implemented on a fixed in-sample testing window W_T , which consists of 3000 observations; this follows advice by Danielsson (2011) which suggests that this testing window W_T should cover at least 4 years of data, or approximately 1000 observations, for a reliable statistical analysis. Specifically, we use:

- The time period from 8 December 1997 to 9 November 2009 for the Dow Jones,
- The time period from 13 August 1997 to 16 July 2009 for the Nasdaq,
- The time period from 29 May 1997 to 12 August 2009 for the Nikkei,
- The time period from 28 September 2002 to 14 December 2010 for the JPY/GBP exchange rate.

This allows us to focus on extreme VaR estimation around the 2007-2008 financial crisis, of which a consequence was a succession of extremely large negative log-returns in a very short timeframe.

This should be considered a challenging problem.

In each case, we implement the three methods on these 3000 observations. The HS and UGH method work directly on the series X_t , without filtering, the estimate then being $\bar{q}_\tau(X_{t+1} | \mathcal{F}_t) = \bar{q}_\tau(X)$, where $\bar{q}_\tau(X)$ is the empirical τ th quantile of the data for the HS method and, for the UGH method, $\bar{q}_\tau(X) = \hat{q}_\tau(X)$ is obtained as in Section 2.2 with the X_t in place of the \hat{Z}_t . By contrast, the GARCH-N, GARCH- t , GARCH-EVT and GARCH-UGH methods filter the data using an AR(1)-GARCH(1,1) model $X_t = \mu_t + \sigma_t Z_t$ with a Gaussian QMLE, and then estimate $q_\tau(Z)$ on the basis of the residuals obtained from this filtering with an approach specific to each method, before obtaining the final extreme VaR estimate by combining the AR(1)-GARCH(1,1) estimates and the estimate of $q_\tau(Z)$. In these four methods, the difference lies in how $q_\tau(Z)$ is estimated. In addition, we calculate another version of the GARCH-UGH estimate where the estimator $\hat{\rho}_{k_p}$ is replaced throughout by the constant -1 , as mentioned in Section 2.2. If this other version has a number of VaR violations closer to the expected number of violations (which is known and equal to $3000(1 - \tau)$ where τ is the VaR level), we retain this version.

Results from Tables 2-5 indicate that, on the basis of in-sample validation and compared to the other two methods geared towards extreme value estimation (GARCH-EVT and UGH), the proposed GARCH-UGH approach is the most successful for estimating one-step ahead extreme VaRs that satisfy both unconditional and conditional coverage properties. Across all samples and in terms of number of VaR violations only, in 46 out of 60 cases our GARCH-UGH approach is closest to the mark, with the conventional GARCH-EVT performing worst overall. In addition, although the unfiltered UGH estimate is somewhat reasonable in terms of number of VaR violations, it is not appropriate because it lacks responsiveness to the time-varying volatility and volatility clustering: Figure 2a illustrates that the non-dynamic nature of the UGH estimate leaves it unable to respond immediately to high volatility, and VaR violations tend to cluster. By contrast, the conditional VaR estimates obtained by our GARCH-UGH approach (Figure 2b) clearly respond to the changing volatility with no clustering of VaR violations, while bias reduction results in closer numbers of VaR violations to the expected numbers than with the conventional GARCH-EVT. Numerically, the GARCH-UGH method never fails either the Kupiec or Christoffersen tests, whereas the GARCH-EVT method fails 7 and 5 times out of 60 cases, respectively. The bias correction at the extreme value step appears to be very effective for the accurate estimation of one-step ahead dynamic extreme VaRs. It leads to results that seem less sensitive to the choice of sample fraction k than the conventional GARCH-EVT method: see Tables 2-5, where results appear to be consistently good across a large range of values of k .

In addition, Table 6 shows the superiority of the GARCH-UGH approach when it is compared with the basic estimation methods (HS, GARCH-N and GARCH- t) that are commonly used by practitioners in financial risk management. Note that the number of VaR violations for GARCH-UGH shown in the Table 6 is when the optimal (according to Tables 2-5) sample fraction is chosen from 5% to 25% for the estimation of dynamic extreme VaR. For all cases, HS provides the same number of VaR violations as theoretically expected ones. This is because the nonparametric HS method gives the (length of testing window $\times (1 - \tau)$)th ordered value in the sample as the VaR at quantile level τ from the non-updated ordered observations, which always ends up producing the same number of VaR violations as theoretically expected. Hence, in-sample HS is trivial and we exclude it from the comparison for in-sample estimation; we will see in the out-of-sample backtestings that HS performs worse out of all estimation methods. Across all samples, in 11 out of 12 cases our GARCH-UGH approach is closest to the mark and never fails any tests, with the GARCH-N approach performing worse since it cannot capture heavy tails. The GARCH-N and GARCH- t methods are not reliable approaches for the estimation of dynamic extreme VaR because GARCH-N fails to pass the Kupiec and Christoffersen tests 6 and 5 times out of 12 cases, and GARCH- t fails 4 and 3 times respectively.

3.3. Out-of-sample dynamic extreme VaR estimation and backtesting

We now focus on the out-of-sample estimation (that is, prediction) of one-step ahead VaR via the same six approaches, again at level $\tau \in \{0.99, 0.995, 0.999\}$. We consider the following samples of data:

- The time period from 23 December 1993 to 9 November 2009 for the Dow Jones,
- The time period from 30 August 1993 to 16 July 2009 for the Nasdaq,
- The time period from 14 May 1993 to 12 August 2009 for the Nikkei,
- The time period from 2 January 2000 to 14 December 2010 for the JPY/GBP exchange rate.

In order to carry out this out-of-sample backtest, we adopt a rolling window estimation approach. Specifically, we first fix a testing window W_T in each case, which corresponds to the periods of time considered in our in-sample evaluation (8 December 1997 to 9 November 2009 for the Dow Jones, 13 August 1997 to 16 July 2009 for the Nasdaq, 29 May 1997 to 12 August 2009 for the Nikkei, 28 September 2002 to 14 December 2010 for the JPY/GBP exchange rate). At each time t in this testing window W_T , we use a window of length W_E of prior information in order to predict the conditional VaR on time $t+1$ (with parameter estimates updated when the estimation window changes), which is then compared to the observed log-return on day $t+1$. Various choices of W_E have been made in the literature: here we choose $W_E = 1000$ as in McNeil and Frey (2000), corresponding to approximately four years of model calibration for each prediction with stock market data, and three years with exchange rate data. Regarding the use of the GARCH-UGH method specifically, we retain the implementation suggested by the results of in-sample backtesting. In other words, on a given data set and for a given value of k , if we observed during in-sample backtesting that the choice $\hat{\rho}_{k\rho} = -1$ performed better, then we retain this choice for out-of-sample estimation; otherwise, we estimate ρ as indicated in Section 2.2.

Tables 7-10 gather the numerical results for the comparison between the GARCH-EVT, GARCH-UGH and UGH methods. It can be seen that again, the suggested GARCH-UGH approach appears to be best overall. In 47 out of 60 cases, the GARCH-UGH approach yields the closest number of VaR violations to the theoretically expected numbers, while the unfiltered UGH method fares worst. Based on the Kupiec test, the GARCH-UGH approach fails twice, whereas the GARCH-EVT and UGH fail 6 and 49 times out of 60 cases, respectively. On one occasion GARCH-UGH fails the Christoffersen test, while the GARCH-EVT and UGH methods fail 0 and 43 times out of 60 cases. The GARCH-UGH typically performs better than other approaches except possibly when the top 5% and 10% of observations are used (for the choice of k); this is because the bias is not the dominating term in the bias-variance tradeoff when k is small. Table 11 also supports the use of GARCH-UGH approach for the estimation of dynamic extreme VaR by outperforming the basic HS, GARCH-N and GARCH- t estimation methods. In 12 out of 12 cases our GARCH-UGH approach (with optimal sample fraction according to Tables 7-10) is closest to the mark. It also never fails either of the Kupiec and Christoffersen tests, while HS fails 8 and 7 times, GARCH-N fails 10 and 10 times, and GARCH- t fails 3 and 2 times out of 12 cases, respectively. The corresponding plots of out-of-sample backtesting are shown in Figures 3-6 with the corresponding 95% asymptotic Gaussian confidence intervals corresponding to the GARCH-UGH estimation method in Figure 7: the confidence interval is given by

$$\left[\hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{q}_{1-p}(Z) \left(1 \pm \frac{1.96}{\sqrt{k}/\log(\frac{k}{np})} \times \sqrt{\frac{\hat{\gamma}_{k,k\rho}^2}{\hat{\rho}_{k\rho}^2} (\hat{\rho}_{k\rho}^2 + (1 - \hat{\rho}_{k\rho})^2)} \right) \right].$$

It is clearly seen that the GARCH-UGH and GARCH-EVT estimates have the same dynamics, with the bias correction shifting the estimate upwards or downwards depending on the rolling estimation window.

4. Discussion

In this paper we introduce an extension of the two-step GARCH-EVT approach from McNeil and Frey (2000) for extreme VaR estimation, based on a semiparametric bias-reduced extreme quantile estimator from Chavez-Demoulin and Guillou (2018) and de Haan *et al.* (2016). This differs from the other papers published in the econometric literature by introducing a finite-sample improvement at the extreme value step, rather than using a more complicated filter than the AR(1)-GARCH(1,1) filter. We conclude from our empirical analysis that the proposed GARCH-UGH approach provides better one-step ahead dynamic extreme VaR estimates for financial time series than the benchmark conventional GARCH-EVT approach of McNeil and Frey (2000) and the other basic estimation approaches we have tested based on historical simulation or traditional fully parametric models. This can be seen from both the in-sample and the out-of-sample estimations at several quantile levels τ , including the very high $\tau = 0.999$ corresponding to a 99.9% VaR, and a large range of sample fractions k , due to the effect of the bias correction. Let us also point out that the GARCH-UGH method is carried out using an automatic recipe for the estimation of the extreme value index and extreme quantile, making it computationally cheap.

We highlight three possible directions for further investigations. The first one is that one could replace the AR(1)-GARCH(1,1) filter by a more sophisticated filter. Which filter should be used is not obvious: one could think about replacing the AR(1) part by an ARMA(p, q) part, or the GARCH(1,1) part by a GARCH(p, q) part (or a more complicated asymmetric version), or both. This may make it possible to even better account for the volatility dynamics, whose accurate estimation and prediction are key. The second one is the extension of our GARCH-UGH approach to the estimation of the multiple-step ahead conditional extreme VaR. This is important, because certain regulations such as those advocated by Basel Committee on Banking Supervision (2009) require the estimation of the 10-day ahead VaR at the 99% confidence level, rather than merely the one-step ahead VaR. This is a challenging problem: McNeil and Frey (2000) tackles this question using a bootstrap methodology, but bootstrapping with heavy tails is known to be very difficult to calibrate, especially in the extreme value setup we consider here. The development of an adaptation of the GARCH-UGH method to the multiple-step ahead setup which stays computationally manageable and accurate is well beyond the scope of the current paper. The third and final perspective is the estimation of alternative dynamic risk measures as a way of solving the two drawbacks of VaR that we highlighted in the previous paragraph. One candidate will be the expectile risk measure (see Newey and Powell 1987 for the original definition of expectiles in a regression context), which takes into account both the frequency of extreme observations and their magnitude, and is also shown to be a coherent and elicitable risk measure in Ziegel (2016). The use of expectiles has recently received substantial attention from the perspective of risk management as an alternative tool for quantifying tail risk (see for example Daouia *et al.* 2018, 2020), but the case of dynamic estimation of extreme expectiles in a financial time series context has not been considered yet. Of course, there exists no universally preferred risk measure: the expectile only has an implicit formulation in general, and is more difficult to interpret than the VaR. The development of a GARCH-UGH-based method for the estimation of dynamic extreme expectiles will thus be an interesting complement to the present paper.

Acknowledgments

The authors thank the Editors and an anonymous referee for their valuable suggestions and comments which led to an improved version of this paper. The first author gratefully acknowledges the financial support from the SOKENDAI (Graduate University for Advanced Studies) under SOKENDAI Student Dispatch Program grant, and the Research Fellowships of Japan Society for the Promotion of Science for Young Scientists (Project number: 20J15188). Y. Kawasaki is supported by JSPS Grant-in-Aid for Scientific Research (18H00836, 19K01597, 20H01502) and by the

ISM Cooperative Research Program 2020-ISMCRP-2038. G. Stupfler is supported by the French National Research Agency under the grant ANR-19-CE40-0013 (ExtremReg project), and he also acknowledges support from an AXA Research Fund Award on “Mitigating risk in the wake of the COVID-19 pandemic”.

References

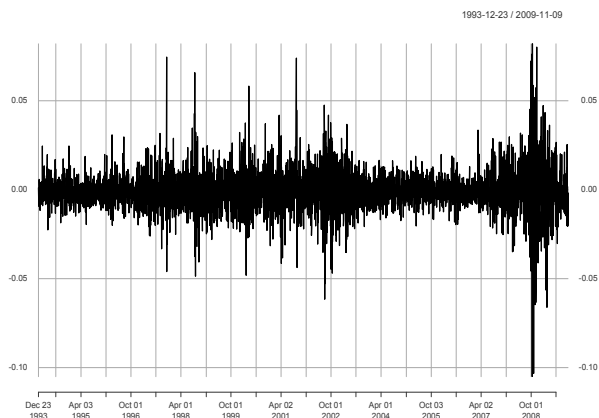
- Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., Coherent Measures of Risk. *Mathematical Finance*, 1999, **9**, 203–228.
- Basel Committee on Banking Supervision, Revisions to the Basel II market risk framework. <https://www.bis.org/publ/bcbs265.pdf> Accessed: 2020-10-14, 2009.
- Basel Committee on Banking Supervision, Fundamental review of the trading book: A revised market risk framework. <https://www.bis.org/publ/bcbs265.pdf> Accessed: 2020-10-14, 2013.
- Bee, M., Dupuis, D. and Trapin, L., Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance*, 2016, **36**, 86–99.
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J., *Statistics of Extremes: Theory and Applications*, 2004, Wiley.
- Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 1986, **31**, 307–327.
- Byström, H., Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 2004, **13**, 133–152.
- Cai, J.J., de Haan, L. and Zhou, C., Bias correction in extreme value statistics with index around zero. *Extremes*, 2013, **16**, 173–201.
- Chavez-Demoulin, V., Davison, A. and McNeil, A., Estimating Value-at-Risk: A point process approach. *Quantitative Finance*, 2005, **5**, 227–234.
- Chavez-Demoulin, V. and Guillou, A., Extreme quantile estimator for β -mixing time series and applications. *Insurance: Mathematics and Economics*, 2018, **83**, 59–74.
- Christoffersen, P., Evaluating interval forecasts. *International Economic Review*, 1998, **39**, 841–862.
- Danielsson, J., *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*, 2011, Wiley.
- Daouia, A., Girard, S. and Stupfler, G., Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society: Series B*, 2018, **80**, 263–292.
- Daouia, A., Girard, S. and Stupfler, G., Tail expectile process and risk assessment. *Bernoulli*, 2020, **26**, 531–556.
- de Haan, L. and Ferreira, A., *Extreme Value Theory An Introduction*, 2006, Springer Series in Operations Research and Financial Engineering. New York: Springer.
- de Haan, L., Mercadier, C. and Zhou, C., Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics*, 2016, **20**, 321–354.
- Deng, K. and Qiu, J., Backtesting expected shortfall and beyond. *Quantitative Finance*, 2021, **21**, 1109–1125.
- Embrechts, P., Klüppelberg, C. and Mikosch, T., *Modelling Extremal Events for Insurance and Finance*, 1997, Springer.
- Ergen, I., Two-step methods in VaR prediction and the importance of fat tails. *Quantitative Finance*, 2015, **15**, 1013–1030.
- Fernandez, V., Risk management under extreme events. *International Review of Financial Analysis*, 2005, **14**, 113–148.
- Francq, C. and Zakoian, J.M., Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes. *Bernoulli*, 2004, **10**, 605–637.
- Francq, C. and Zakoian, J.M., *GARCH Models: Structure, Statistical Inference and Financial Applications*, 2010, John Wiley & Sons.
- Furió, D. and Climent, F., Extreme value theory versus traditional GARCH approaches applied to financial data: a comparative evaluation. *Quantitative Finance*, 2013, **13**, 45–63.
- Galanos, A., rugarch: Univariate GARCH Models. R package version 1.4-1, 2019.
- Gneiting, T., Making and evaluating point forecasts. *Journal of the American Statistical Association*, 2011, **106**, 746–762.

- Goegebeur, Y., Beirlant, J. and de Wet, T., Kernel estimators for the second order parameter in extreme value statistics. *Journal of Statistical Planning and Inference*, 2010, **140**, 2632–2652.
- Gomes, M.I., Pestana, D. and Caeiro, F., A note on the asymptotic variance at optimal levels of a bias-corrected Hill estimator. *Statistics & Probability Letters*, 2009, **79**, 295–303.
- Gomes, M., de Haan, L. and Peng, L., Semi-parametric Estimation of the Second Order Parameter in Statistics of Extremes. *Extremes*, 2002, **5**, 387–414.
- Gomes, M., Martins, M. and Neves, M., Alternatives to a Semi-Parametric Estimator of Parameters of Rare Events - The Jackknife Methodology. *Extremes*, 2000, **3**, 207–229.
- Hill, B.M., A simple general approach to inference about the tails of a distribution. *The Annals of Statistics*, 1975, **3**, 1163–1174.
- Hill, J.B., Tail index estimation for a filtered dependent time series. *Statistica Sinica*, 2015, **25**, 609–629.
- Hofert, M. and Hornik, K., qrmdata: Data Sets for Quantitative Risk Management Practice. R package version 2016-01-03-1, 2016.
- Jalal, A. and Rockinger, M., Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data. *Journal of Empirical Finance*, 2008, **15**, 868–877.
- Kupiec, P., Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 1995, **3**, 73–84.
- McNeil, A. and Frey, R., Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 2000, **7**, 271–300.
- McNeil, A., Frey, R. and Embrechts, P., *Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition*, 2015, Princeton University Press.
- Newey, W.K. and Powell, J.L., Asymmetric Least Squares Estimation and Testing. *Econometrica*, 1987, **55**, 819–847.
- Reiss, R. and Thomas, M., *Statistical Analysis of Extreme Values*, 2007, Birkhauser Verlag.
- Resnick, S., *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, 2007, Springer.
- Weissman, I., Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 1978, **73**, 812–815.
- Yi, Y., Feng, X. and Huang, Z., Estimation of extreme value-at-risk: An EVT approach for quantile GARCH model. *Econometric Letters*, 2014, **124**, 378–381.
- Youssef, M., Belkacem, L. and Mokni, K., Value-at-Risk estimation of energy commodities: A long-memory GARCH-EVT approach. *Energy Economics*, 2015, **51**, 99–110.
- Zhao, L.T., Liu, K., Duan, X.L. and Li, M.F., Oil price risk evaluation using a novel hybrid model based on time-varying long memory. *Energy Economics*, 2019, **81**, 70–78.
- Ziegel, J., Coherence and elicibility. *Mathematical Finance*, 2016, **24**, 901–918.

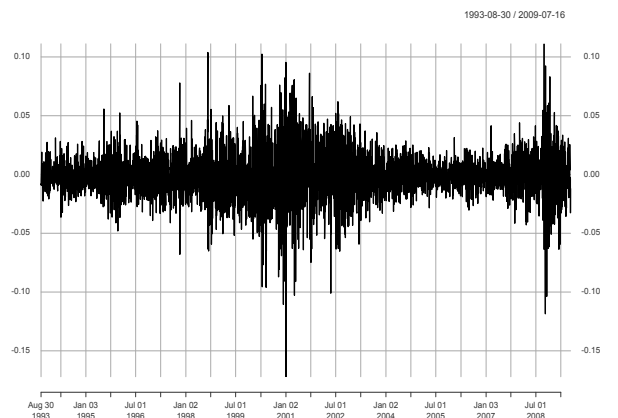
Table 1.: Summary of descriptive statistics and basic statistical tests for daily negative log-returns on DJ, NASDAQ, NIKKEI and JPY/GBP.

	DJ	NASDAQ	NIKKEI	JPY/GBP
Sample size	4000	4000	4000	4000
Mean	-0.000250	-0.000355	-0.000169	-0.0000557
Median	-0.000460	-0.00123	-0.0000177	0
Maximum	0.0820	0.111	0.121	0.0600
Minimum	-0.105	-0.172	-0.132	-0.0640
Standard deviation	0.0119	0.0203	0.0155	0.00626
Skewness	0.117	-0.110	0.175	-0.586
Kurtosis	8.096	4.469	5.579	10.931
J-B test	10933* (0.0000)	3337.3* (0.0000)	5207.9* (0.0000)	2014.6* (0.0000)
$Q(1)$	13.159* (0.000)	12.098* (0.001)	6.680* (0.010)	128.68* (0.000)
$Q(5)$	37.723* (0.000)	37.62* (0.000)	14.429* (0.013)	146.47* (0.000)
$Q(10)$	50.388* (0.000)	42.192* (0.000)	23.023* (0.011)	150.37* (0.000)
$Q^2(1)$	131.54* (0.000)	207.21* (0.000)	248.32* (0.000)	275.36* (0.000)
$Q^2(10)$	2613.3* (0.000)	1907.5* (0.000)	3183.4* (0.000)	1650.4* (0.000)
ADF test	-15.782**	-14.794**	-15.967**	-16.415**

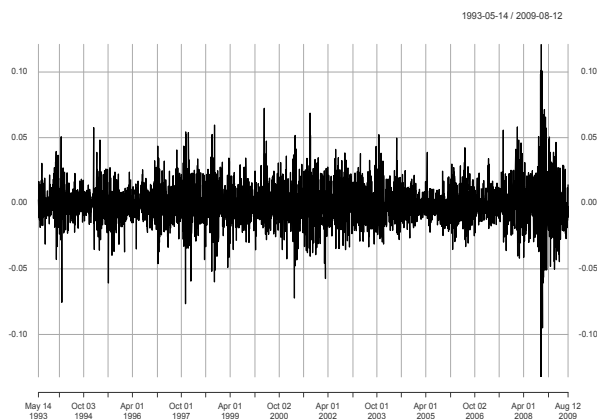
Notes: A kurtosis greater than 3 indicates that the dataset has heavier tails than a normal distribution. J-B stands for the Jarque-Bera test, $Q(n)$ and $Q^2(n)$ are the Ljung-Box tests for autocorrelation at lags n in the negative log-return series and squared negative log-returns, respectively. The ADF test is the augmented Dickey-Fuller stationarity test statistic without trend. The p -values are given between brackets. **, * denote significance at 1% and 5% levels, respectively.



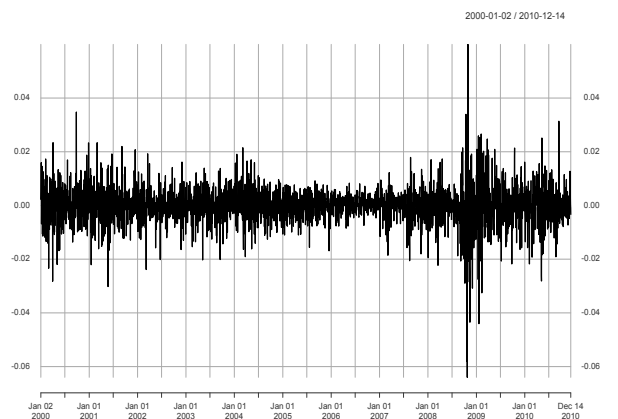
(a) DJ (23/12/1993 to 09/11/2009)



(b) NASDAQ (30/08/1993 to 16/07/2009)



(c) NIKKEI (14/05/1993 to 12/08/2009)



(d) JPY/GBP (02/01/2000 to 14/12/2010)

Figure 1.: Daily negative log-returns of four financial time series: DJ, NASDAQ, NIKKEI and JPY/GBP.

Table 2.: In-sample evaluations of one-step ahead conditional VaR estimates from 8 December 1997 to 9 November 2009 at different quantile levels for the negative log-returns of DJ index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
% of top obs. used	5%	10%	15%	20%	25%
DJ:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	4	5	2	2	3
	(0.583, 0.855)	(0.292, 0.569)	(0.538, 0.826)	(0.538, 0.826)	(1.000, 0.997)
GARCH-UGH	2	2	2	2	2
	(0.538, 0.826)	(0.538, 0.826)	(0.538, 0.826)	(0.538, 0.826)	(0.538, 0.826)
GARCH-EVT	2	2	2	2	4
	(0.538, 0.826)	(0.538, 0.826)	(0.538, 0.826)	(0.538, 0.826)	(0.583, 0.855)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	18	18	16	18	20
	(0.452, 0.012)	(0.452, 0.012)	(0.798, 0.009)	(0.452, 0.012)	(0.218, 0.011)
GARCH-UGH	15	14	14	15	15
	(1.000, 0.927)	(0.793, 0.905)	(0.793, 0.905)	(1.000, 0.927)	(1.000, 0.927)
GARCH-EVT	13	13	13	13	13
	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	34	34	34	36	39
	(0.472, 0.018)	(0.472, 0.018)	(0.472, 0.018)	(0.286, 0.018)	(0.114, 0.014)
GARCH-UGH	27	28	29	31	33
	(0.576, 0.669)	(0.711, 0.717)	(0.854, 0.741)	(0.855, 0.711)	(0.588, 0.598)
GARCH-EVT	23	23	22	22	20
	(0.180, 0.341)	(0.180, 0.341)	(0.123, 0.259)	(0.123, 0.259)	(0.050, 0.130)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

Table 3.: In-sample evaluations of one-step ahead conditional VaR estimates from 13 August 1997 to 16 July 2009 at different quantile levels for the negative log-returns of NASDAQ index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
% of top obs. used	5%	10%	15%	20%	25%
NASDAQ:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	3	1	1	1	1
	(1.000, 0.997)	(0.179, 0.406)	(0.179, 0.406)	(0.179, 0.406)	(0.179, 0.406)
GARCH-UGH	4	4	4	4	2
	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.538, 0.826)
GARCH-EVT	4	4	4	4	4
	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	21	21	21	19	21
	(0.143, 0.295)	(0.143, 0.295)	(0.143, 0.295)	(0.320, 0.541)	(0.143, 0.295)
GARCH-UGH	14	14	14	14	13
	(0.793, 0.905)	(0.793, 0.905)	(0.793, 0.905)	(0.793, 0.905)	(0.596, 0.821)
GARCH-EVT	13	13	10	10	10
	(0.596, 0.821)	(0.596, 0.821)	(0.168, 0.374)	(0.168, 0.374)	(0.168, 0.374)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	32	33	33	35	37
	(0.717, 0.609)	(0.588, 0.135)	(0.588, 0.135)	(0.371, 0.127)	(0.215, 0.106)
GARCH-UGH	23	23	23	25	25
	(0.180, 0.341)	(0.180, 0.341)	(0.180, 0.341)	(0.345, 0.519)	(0.345, 0.519)
GARCH-EVT	22	17	16	16	16
	(0.123, 0.259)	(0.009, 0.031)	(0.005, 0.017)	(0.005, 0.017)	(0.005, 0.017)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

Table 4.: In-sample evaluations of one-step ahead conditional VaR estimates from 29 May 1997 to 12 August 2009 at different quantile levels for the negative log-returns of NIKKEI index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
% of top obs. used	5%	10%	15%	20%	25%
<i>NIKKEI:</i>					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	4	4	4	4	1
	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.583, 0.855)	(0.179, 0.406)
GARCH-UGH	4	2	4	4	1
	(0.583, 0.885)	(0.538, 0.826)	(0.583, 0.885)	(0.583, 0.885)	(0.179, 0.406)
GARCH-EVT	5	5	5	5	5
	(0.292, 0.569)	(0.292, 0.569)	(0.292, 0.569)	(0.292, 0.569)	(0.292, 0.569)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	15	15	17	18	21
	(1.000, 0.178)	(1.000, 0.178)	(0.612, 0.199)	(0.452, 0.190)	(0.143, 0.114)
GARCH-UGH	13	13	13	13	12
	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)	(0.421, 0.689)
GARCH-EVT	13	12	12	12	12
	(0.596, 0.821)	(0.421, 0.689)	(0.421, 0.689)	(0.421, 0.689)	(0.421, 0.689)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	32	32	34	36	38
	(0.717, 0.609)	(0.717, 0.609)	(0.472, 0.562)	(0.286, 0.427)	(0.159, 0.297)
GARCH-UGH	26	25	26	31	28
	(0.453, 0.601)	(0.345, 0.519)	(0.453, 0.601)	(0.855, 0.711)	(0.711, 0.666)
GARCH-EVT	25	24	21	19	18
	(0.345, 0.287)	(0.254, 0.430)	(0.081, 0.188)	(0.030, 0.085)	(0.017, 0.049)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

Table 5.: In-sample evaluations of one-step ahead conditional VaR estimates from 28 September 2002 to 14 December 2010 at different quantile levels for the negative log-returns of JPY/GBP exchange rate by means of the number of VaR violations, unconditional and conditional coverage tests.

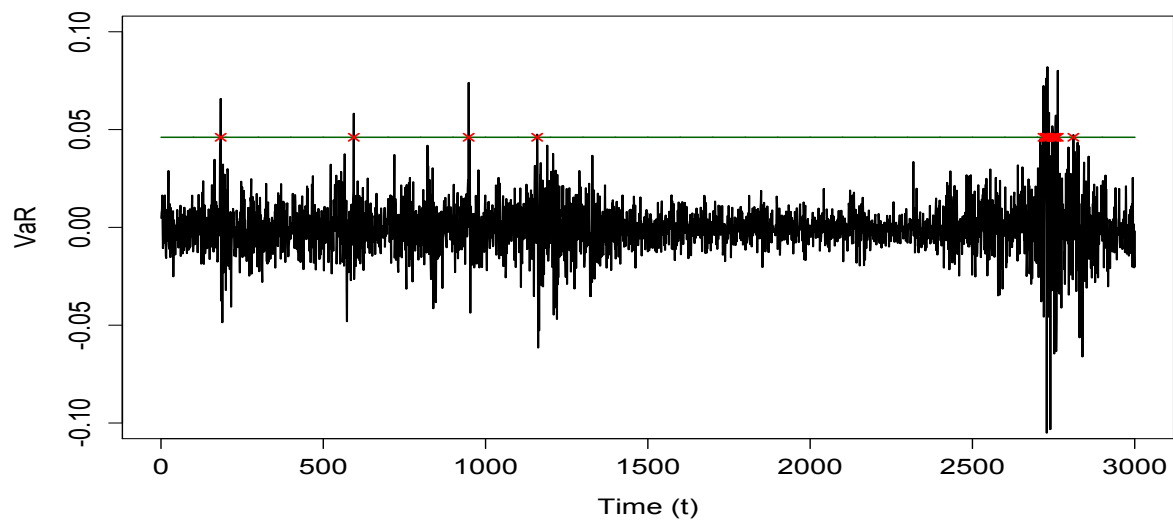
Testing window	3000				
% of top obs. used	5%	10%	15%	20%	25%
<i>JPY/GBP:</i>					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	2	2	1	1	1
	(0.538, 0.826)	(0.538, 0.826)	(0.179, 0.406)	(0.179, 0.406)	(0.179, 0.406)
GARCH-UGH	3	2	3	2	2
	(1.000, 0.997)	(0.538, 0.826)	(1.000, 0.997)	(0.538, 0.826)	(0.538, 0.826)
GARCH-EVT	3	3	3	3	3
	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	16	17	16	18	28
	(0.798, 0.195)	(0.612, 0.199)	(0.798, 0.195)	(0.452, 0.190)	(0.003, 0.000)
GARCH-UGH	16	14	14	14	16
	(0.798, 0.888)	(0.793, 0.905)	(0.793, 0.905)	(0.793, 0.905)	(0.798, 0.888)
GARCH-EVT	11	11	11	11	10
	(0.277, 0.532)	(0.277, 0.532)	(0.277, 0.532)	(0.277, 0.532)	(0.168, 0.374)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	38	40	41	41	46
	(0.159, 0.002)	(0.081, 0.002)	(0.056, 0.001)	(0.056, 0.001)	(0.006, 0.001)
GARCH-UGH	31	32	31	29	22
	(0.855, 0.612)	(0.717, 0.609)	(0.855, 0.612)	(0.854, 0.556)	(0.123, 0.259)
GARCH-EVT	29	29	28	24	22
	(0.854, 0.556)	(0.854, 0.556)	(0.711, 0.501)	(0.254, 0.430)	(0.123, 0.259)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

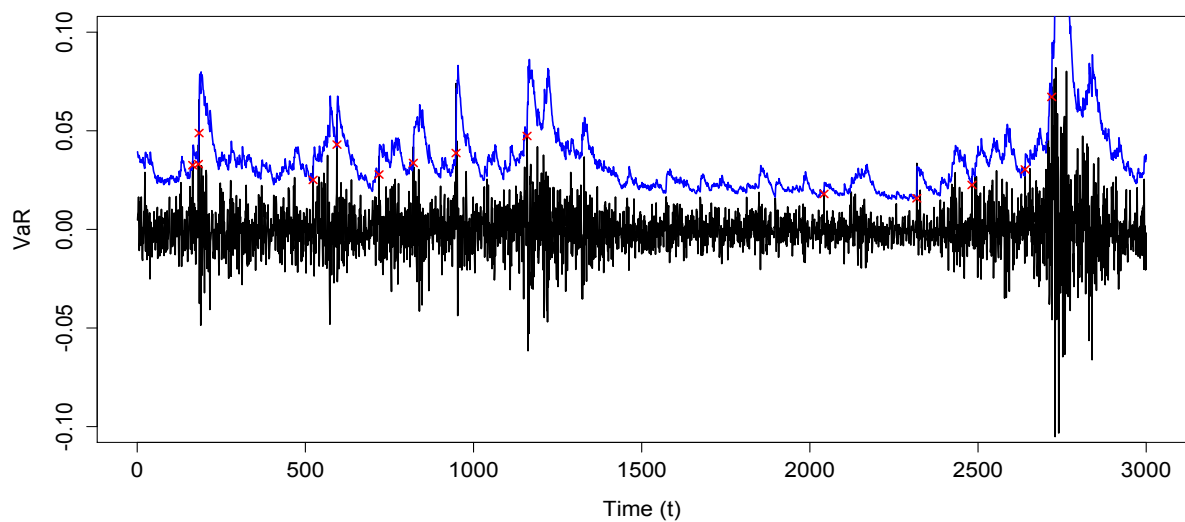
Table 6.: In-sample evaluations of one-step ahead conditional VaR estimates at different quantile levels for the negative log-returns of DJ, NASDAQ, NIKKEI indices and JPY/GBP exchange rate (time period given in Section 3.2) by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	DJ 3000	NASDAQ 3000	NIKKEI 3000	JPY/GBP 3000
<i>0.999 Quantile</i>				
Expected	3	3	3	3
HS	3	3	3	3
	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)
GARCH-N	13	9	11	7
	(0.000, 0.000)	(0.005, 0.020)	(0.000, 0.002)	(0.049, 0.142)
GARCH- <i>t</i>	2	4	2	1
	(0.538, 0.826)	(0.583, 0.855)	(0.538, 0.826)	(0.179, 0.406)
GARCH-UGH	2	4	4	3
	(0.538, 0.826)	(0.583, 0.855)	(0.583, 0.855)	(1.000, 0.997)
<i>0.995 Quantile</i>				
Expected	15	15	15	15
HS	15	15	15	15
	(1.000, 0.007)	(1.000, 0.923)	(1.000, 0.178)	(1.000, 0.178)
GARCH-N	28	16	25	20
	(0.003, 0.008)	(0.798, 0.888)	(0.018, 0.005)	(0.218, 0.410)
GARCH- <i>t</i>	13	10	12	1
	(0.596, 0.821)	(0.168, 0.374)	(0.421, 0.689)	(0.000, 0.000)
GARCH-UGH	15	14	13	16
	(1.000, 0.927)	(0.793, 0.905)	(0.596, 0.821)	(0.798, 0.888)
<i>0.99 Quantile</i>				
Expected	30	30	30	30
HS	30	30	30	30
	(1.000, 0.112)	(1.000, 0.594)	(1.000, 0.594)	(1.000, 0.012)
GARCH-N	43	27	41	38
	(0.025, 0.044)	(0.576, 0.669)	(0.056, 0.091)	(0.159, 0.297)
GARCH- <i>t</i>	20	16	18	3
	(0.051, 0.130)	(0.005, 0.017)	(0.017, 0.053)	(0.000, 0.000)
GARCH-UGH	29	25	31	31
	(0.854, 0.741)	(0.345, 0.519)	(0.855, 0.711)	(0.855, 0.612)

Notes: The closest number of VaR violations to the theoretically expected number is highlighted in bold, excluding historical simulation (HS), see Section 3.2. The number of VaR violations for GARCH-UGH is reported when the optimal sample fraction is selected according to Tables 2-5. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at 5% significance level are given in brackets in order.



(a) UGH approach



(b) GARCH-UGH approach

Figure 2.: Twelve years (8 December 1997 to 9 November 2009) of in-sample backtesting of the DJ index, and 99.5%-VaR violations by (a) the UGH approach and (b) the GARCH-UGH approach when the top 15% of observations are used for the estimation. Red cross marks denote the VaR violations.

Table 7.: Out-of-sample evaluations of one-step ahead conditional VaR estimates from 8 December 1997 to 9 November 2009 at different quantile levels for the negative log-returns of DJ index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
Estimation window	1000				
% of top obs. used	5%	10%	15%	20%	25%
DJ:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	10	9	9	7	6
	(0.001, 0.006)	(0.005, 0.020)	(0.005, 0.020)	(0.049, 0.142)	(0.128, 0.310)
GARCH-UGH	3	3	3	3	3
	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)
GARCH-EVT	3	4	4	4	4
	(1.000, 0.997)	(0.583, 0.885)	(0.583, 0.885)	(0.583, 0.885)	(0.583, 0.885)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	40	40	40	36	29
	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.001, 0.001)
GARCH-UGH	19	18	18	16	14
	(0.320, 0.541)	(0.452, 0.676)	(0.452, 0.676)	(0.798, 0.888)	(0.793, 0.905)
GARCH-EVT	19	18	18	17	17
	(0.320, 0.541)	(0.452, 0.676)	(0.452, 0.676)	(0.612, 0.798)	(0.612, 0.798)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	62	64	63	63	61
	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)
GARCH-UGH	33	35	32	31	28
	(0.588, 0.598)	(0.371, 0.433)	(0.717, 0.663)	(0.855, 0.711)	(0.711, 0.717)
GARCH-EVT	33	30	30	28	27
	(0.588, 0.598)	(1.000, 0.738)	(1.000, 0.738)	(0.711, 0.717)	(0.576, 0.669)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

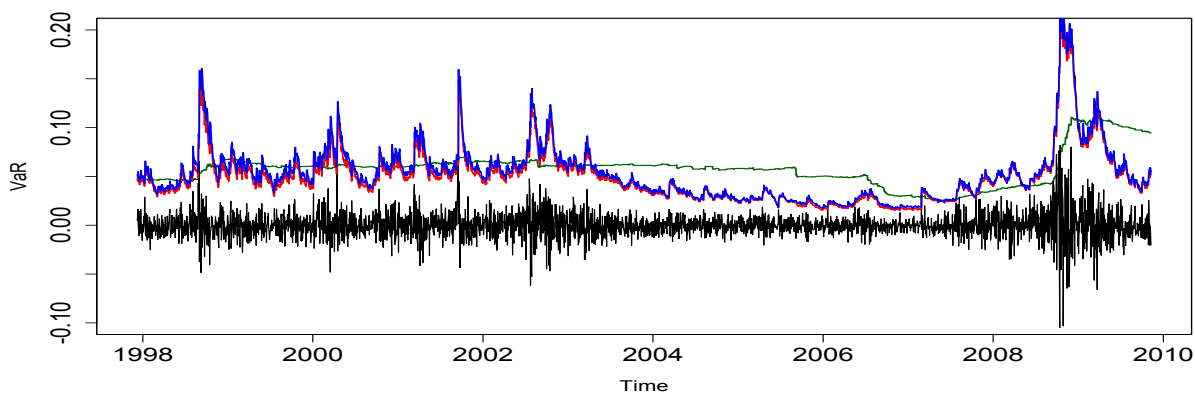


Figure 3.: Out-of-sample backtesting of the DJ index from 8 December 1997 to 9 November 2009, and 99.9%-VaR estimates calculated using rolling estimation windows made of 1000 observations, with k corresponding to the top 15% observations from this window. GARCH-UGH (blue line), GARCH-EVT (red line) and UGH (dark green line) estimates are superimposed on the negative log-returns (black line).

Table 8.: Out-of-sample evaluations of one-step ahead conditional VaR estimates from 13 August 1997 to 16 July 2009 at different quantile levels for the negative log-returns of NASDAQ index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
Estimation window	1000				
% of top obs. used	5%	10%	15%	20%	25%
NASDAQ:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	10	8	7	4	3
	(0.001, 0.006)	(0.017, 0.057)	(0.049, 0.142)	(0.583, 0.855)	(1.000, 0.997)
GARCH-UGH	6	5	5	4	3
	(0.128, 0.370)	(0.292, 0.569)	(0.292, 0.569)	(0.583, 0.855)	(1.000, 0.997)
GARCH-EVT	7	7	7	7	7
	(0.049, 0.142)	(0.049, 0.142)	(0.049, 0.142)	(0.049, 0.142)	(0.049, 0.142)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	39	37	35	36	40
	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)
GARCH-UGH	20	17	15	16	13
	(0.218, 0.410)	(0.612, 0.798)	(1.000, 0.927)	(0.798, 0.888)	(0.596, 0.821)
GARCH-EVT	16	14	13	13	13
	(0.798, 0.888)	(0.793, 0.905)	(0.596, 0.821)	(0.596, 0.821)	(0.596, 0.821)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	74	74	70	65	62
	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)
GARCH-UGH	34	35	31	30	25
	(0.427, 0.544)	(0.371, 0.490)	(0.855, 0.612)	(1.000, 0.594)	(0.345, 0.287)
GARCH-EVT	31	28	28	24	23
	(0.855, 0.612)	(0.711, 0.501)	(0.711, 0.501)	(0.254, 0.430)	(0.180, 0.341)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

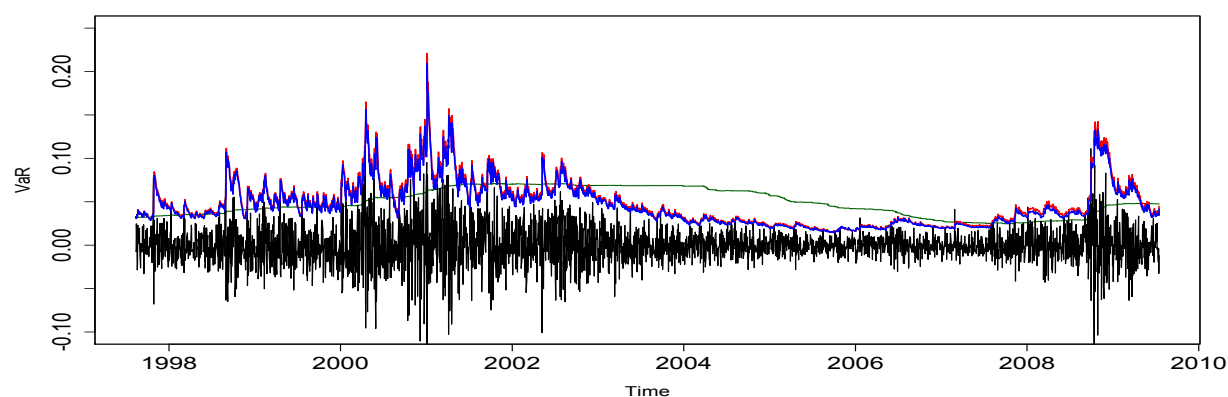


Figure 4.: Out-of-sample backtesting of the NASDAQ index from 13 August 1997 to 16 July 2009, and 99.9%-VaR estimates calculated using rolling estimation windows made of 1000 observations, with k corresponding to the top 20% of observations from this window. GARCH-UGH (blue line), GARCH-EVT (red line) and UGH (dark green line) estimates are superimposed on the negative log-returns (black line).

Table 9.: Out-of-sample evaluations of one-step ahead conditional VaR estimates from 29 May 1997 to 12 August 2009 at different quantile levels for the negative log-returns of NIKKEI index by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
Estimation window	1000				
% of top obs. used	5%	10%	15%	20%	25%
NIKKEI:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	7	6	6	5	5
	(0.049, 0.142)	(0.128, 0.310)	(0.128, 0.310)	(0.292, 0.569)	(0.292, 0.569)
GARCH-UGH	4	3	2	2	1
	(0.583, 0.855)	(1.000, 0.997)	(0.538, 0.826)	(0.538, 0.826)	(0.179, 0.406)
GARCH-EVT	5	4	6	6	6
	(0.292, 0.569)	(0.583, 0.855)	(0.128, 0.310)	(0.128, 0.310)	(0.128, 0.310)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	34	34	34	30	23
	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.000, 0.000)	(0.055, 0.062)
GARCH-UGH	15	15	15	15	12
	(1.000, 0.927)	(1.000, 0.927)	(1.000, 0.927)	(1.000, 0.927)	(0.421, 0.689)
GARCH-EVT	13	14	13	12	12
	(0.596, 0.821)	(0.793, 0.905)	(0.596, 0.821)	(0.421, 0.689)	(0.421, 0.689)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	46	47	46	45	53
	(0.006, 0.011)	(0.004, 0.007)	(0.004, 0.007)	(0.010, 0.015)	(0.000, 0.000)
GARCH-UGH	33	33	33	30	36
	(0.588, 0.598)	(0.588, 0.598)	(0.588, 0.598)	(1.000, 0.738)	(0.286, 0.365)
GARCH-EVT	32	29	27	27	26
	(0.717, 0.663)	(0.854, 0.741)	(0.576, 0.669)	(0.576, 0.669)	(0.453, 0.601)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

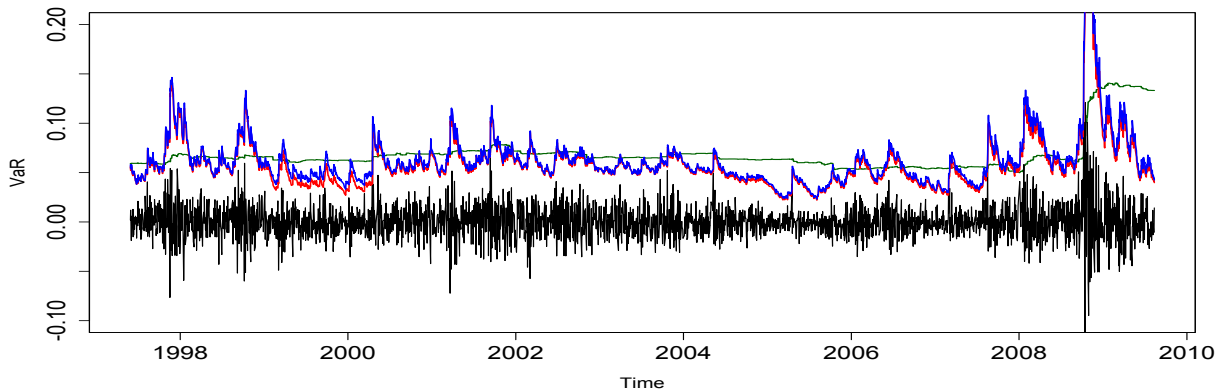


Figure 5.: Out-of-sample backtesting of the NIKKEI index from 29 May 1997 to 12 August 2009, and 99.9%-VaR estimates calculated using rolling estimation windows made of 1000 observations, with k corresponding to the top 10% of observations from this window. GARCH-UGH (blue line), GARCH-EVT (red line) and UGH (dark green line) estimates are superimposed on the negative log-returns (black line).

Table 10.: Out-of-sample evaluations of one-step ahead conditional VaR estimates from 28 September 2002 to 14 December 2010 at different quantile levels for the negative log-returns of JPY/GBP exchange rate by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	3000				
Estimation window	1000				
% of top obs. used	5%	10%	15%	20%	25%
JPY/GBP:					
<i>0.999 Quantile</i>					
Expected	3	3	3	3	3
UGH	7 (0.049, 0.142)	7 (0.049, 0.142)	6 (0.128, 0.310)	4 (0.583, 0.855)	4 (0.583, 0.855)
GARCH-UGH	3 (1.000, 0.997)	2 (0.538, 0.826)	2 (0.538, 0.826)	2 (0.538, 0.826)	2 (0.538, 0.826)
GARCH-EVT	6 (0.128, 0.310)	5 (0.292, 0.569)	5 (0.292, 0.569)	6 (0.128, 0.310)	7 (0.049, 0.142)
<i>0.995 Quantile</i>					
Expected	15	15	15	15	15
UGH	25 (0.018, 0.028)	27 (0.005, 0.010)	27 (0.005, 0.010)	34 (0.000, 0.000)	45 (0.000, 0.000)
GARCH-UGH	21 (0.143, 0.295)	18 (0.452, 0.676)	15 (1.000, 0.927)	14 (0.793, 0.905)	12 (0.421, 0.689)
GARCH-EVT	19 (0.320, 0.541)	19 (0.320, 0.541)	20 (0.219, 0.410)	20 (0.219, 0.410)	20 (0.219, 0.410)
<i>0.99 Quantile</i>					
Expected	30	30	30	30	30
UGH	47 (0.004, 0.000)	56 (0.000, 0.000)	55 (0.000, 0.000)	59 (0.000, 0.000)	67 (0.000, 0.000)
GARCH-UGH	42 (0.038, 0.064)	46 (0.006, 0.012)	40 (0.081, 0.127)	38 (0.159, 0.227)	34 (0.472, 0.523)
GARCH-EVT	38 (0.159, 0.227)	37 (0.215, 0.292)	38 (0.159, 0.227)	38 (0.159, 0.227)	36 (0.286, 0.365)

Notes: The closest numbers of VaR violations to theoretically expected ones are highlighted in bold. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at the 5% significance level are given in brackets in order.

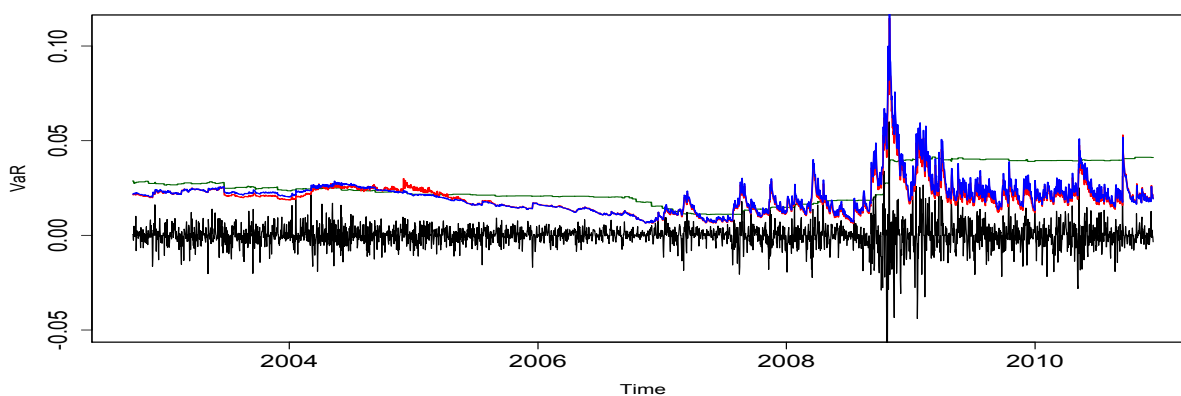


Figure 6.: Out-of-sample backtesting of the JPY/GBP exchange rate from 28 September 2002 to 14 December 2010, and 99.9%-VaR estimates calculated using rolling estimation windows made of 1000 observations, with k corresponding to the top 10% of observations from this window. GARCH-UGH (blue line), GARCH-EVT (red line) and UGH (dark green line) estimates are superimposed on the negative log-returns (black line).

Table 11.: Out-of-sample evaluations of one-step ahead conditional VaR estimates at different quantile levels for the negative log-returns of DJ, NASDAQ, NIKKEI indices and JPY/GBP exchange rate (time period given in Section 3.3) by means of the number of VaR violations, unconditional and conditional coverage tests.

Testing window	DJ	NASDAQ	NIKKEI	JPY/GBP
	3000	3000	3000	3000
Estimation window	1000	1000	1000	1000
<i>0.999 Quantile</i>				
Expected	3	3	3	3
HS	4	5	7	6
	(0.583, 0.855)	(0.292, 0.569)	(0.049, 0.142)	(0.128, 0.310)
GARCH-N	19	11	11	10
	(0.000, 0.000)	(0.000, 0.002)	(0.000, 0.002)	(0.001, 0.006)
GARCH- <i>t</i>	3	7	5	1
	(1.000, 0.997)	(0.049, 0.142)	(0.292, 0.569)	(0.179, 0.406)
GARCH-UGH	3	3	3	3
	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)	(1.000, 0.997)
<i>0.995 Quantile</i>				
Expected	15	15	15	15
HS	36	39	24	21
	(0.000, 0.000)	(0.000, 0.000)	(0.032, 0.042)	(0.143, 0.114)
GARCH-N	34	22	29	29
	(0.000, 0.000)	(0.090, 0.086)	(0.001, 0.004)	(0.001, 0.004)
GARCH- <i>t</i>	17	16	13	1
	(0.612, 0.798)	(0.798, 0.888)	(0.596, 0.821)	(0.000, 0.000)
GARCH-UGH	14	15	15	15
	(0.793, 0.905)	(1.000, 0.927)	(1.000, 0.927)	(1.000, 0.927)
<i>0.99 Quantile</i>				
Expected	30	30	30	30
HS	57	68	44	44
	(0.000, 0.000)	(0.000, 0.000)	(0.016, 0.022)	(0.016, 0.022)
GARCH-N	56	38	44	45
	(0.000, 0.000)	(0.159, 0.297)	(0.016, 0.029)	(0.010, 0.019)
GARCH- <i>t</i>	26	25	20	2
	(0.453, 0.601)	(0.345, 0.287)	(0.051, 0.130)	(0.000, 0.000)
GARCH-UGH	31	30	30	34
	(0.855, 0.711)	(1.000, 0.594)	(1.000, 0.738)	(0.472, 0.523)

Notes: The closest number of VaR violations to the theoretically expected number is highlighted in bold. The number of VaR violations for GARCH-UGH is when the optimal sample fraction is selected according to Tables 7-10. The p-values for the unconditional coverage test by Kupiec (1995) and conditional coverage test by Christoffersen (1998) at 5% significance level are given in brackets in order.

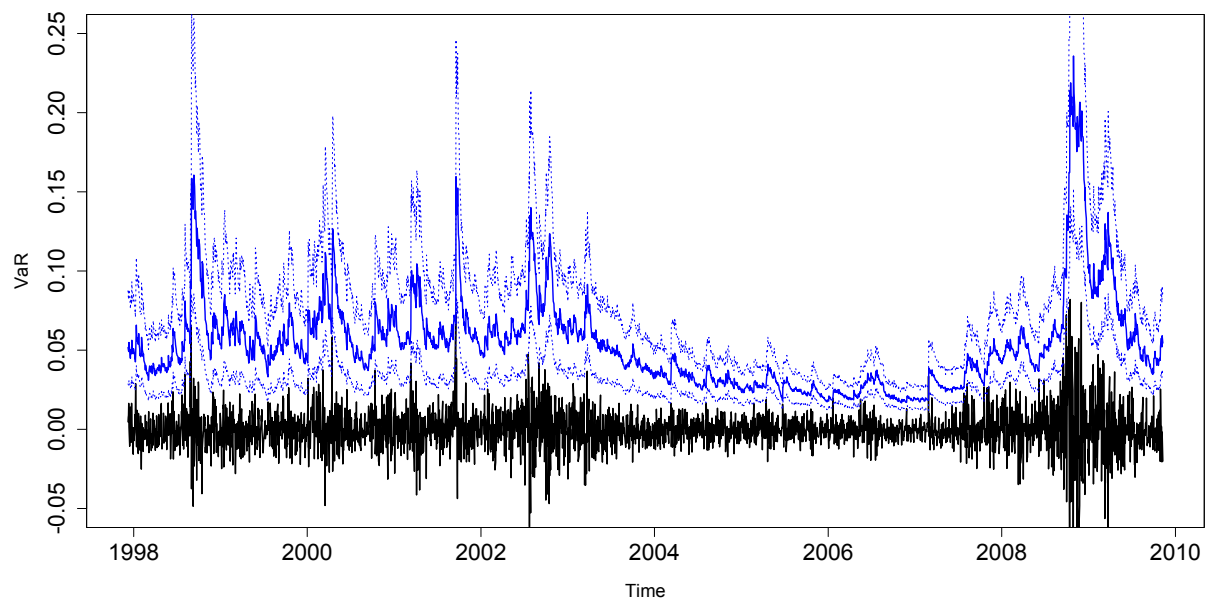


Figure 7.: Out-of-sample backtesting of the DJ index from 8 December 1997 to 9 November 2009, and 99.9%-VaR estimates calculated using rolling estimation windows made of 1000 observations, with k corresponding to the top 15% observations from this window. GARCH-UGH (blue solid line) estimates are superimposed on the negative log-returns (black line) with the 95% asymptotic Gaussian confidence intervals (blue dashed line).