Value-at-Risk Prediction: A Comparison of Alternative Strategies — Source link

Keith Kuester, Stefan Mittnik, Marc S. Paolella

Institutions: Goethe University Frankfurt, Ifo Institute for Economic Research, University of Zurich

Related papers:

- Evaluating Interval Forecasts.

- CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles *

- Generalized autoregressive conditional heteroskedasticity

- Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach

- Techniques for Verifying the Accuracy of Risk Measurement Models

Value-at-Risk Prediction: A Comparison of Alternative Strategies

KEITH KUESTER
   University of Frankfurt

STEFAN MITTNIK
   *University of Munich, Center for Financial Studies,
   and Ifo Institute for Economic Research*

MARC S. PAOLELLA
   University of Zurich

ABSTRACT

Given the growing need for managing financial risk, risk prediction plays an increasing role in banking and finance. In this study we compare the out-of-sample performance of existing methods and some new models for predicting value-at-risk (VaR) in a univariate context. Using more than 30 years of the daily return data on the NASDAQ Composite Index, we find that most approaches perform inadequately, although several models are acceptable under current regulatory assessment rules for model adequacy. A hybrid method, combining a heavy-tailed generalized autoregressive conditionally heteroskedastic (GARCH) filter with an extreme value theory-based approach, performs best overall, closely followed by a variant on a filtered historical simulation, and a new model based on heteroskedastic mixture distributions. Conditional autoregressive VaR (CAViaR) models perform inadequately, though an extension to a particular CAViaR model is shown to outperform the others.

KEYWORDS: empirical finance, extreme value theory, fat tails, GARCH, quantile regression

The market crash in October 1987, recent crises in emerging markets, and disasterous losses resulting from trading activities of institutions—such as Orange County, Long-Term Capital Management Fund, and Metallgesellschaft—have increased the regulatory demand for reliable quantitative risk management tools. See, for example, Gallati (2003:chap. 6) for a set of detailed case studies. The value-at-risk (VaR) concept has emerged as the most prominent measure of downside market risk. It places an upper bound on losses in the sense that these will exceed the VaR threshold with only a small target probability, λ, typically chosen between 1% and 5%. More specifically, conditional on the information given up to time t, the VaR for period $t+h$ of one unit of investment is the negative λ-quantile of the conditional return distribution, that is,

$$\mathrm{VaR}_{t+h}^{\lambda} := -Q_\lambda(r_{t+h}|\mathcal{F}_t) = -\inf_x\{x \in \mathbb{R} : \mathrm{P}(r_{t+h} \le x|\mathcal{F}_t) \ge \lambda\},\ 0 < \lambda < 1, \quad (1)$$

where $Q_\lambda(\cdot)$ denotes the quantile function, r_t is the return on an asset or portfolio in period t, and $\mathcal{F}_t$ represents the information available at date t. We subsequently suppress superscript λ for simplicity.

Regardless of some of its criticisms,[1] regulatory requirements are heavily geared toward VaR.[2] In light of this practical relevance of the VaR concept, the need for reliable VaR estimation and prediction strategies arises. The purpose of this article is to compare alternative approaches for univariate VaR prediction, introduce some new models and provide some guidance for choosing an appropriate strategy.

The work by Bao, Lee, and Saltoglu (2003, 2004) is also concerned with VaR prediction and is a good complement to this article, as they use different data, models, and loss functions. Moreover, their findings, where there is overlap, agree with ours.[3] Other comparison-type studies include Pritsker (1997), although

---

[1] The appropriateness of VaR as a risk measure has been questioned. Evaluating prospects by VaR for varying $\lambda \in (0,1)$ is equivalent to checking for first-order stochastic dominance [as, e.g., implied by the results in Bawa (1978)] and thus does not use the concept of risk aversion to rank prospects. But, by taking just one target probability level, λ, in contrast to using first-order stochastic dominance, any investment will be ranked. Thus, for a specific λ, VaR is a risk measure [cf. Pedersen and Satchell (1998)].

According to the definition in Artzner et al. (1999), VaR fails to be a coherent risk measure. It can lead to Pareto-inferior allocations if agents are risk averse. Further, VaR can fail to appropriately account for portfolio risk diversification [Artzner et al. (1999)].

Further accounts of the problems with VaR can be found in Dowd (2002:sect. 2.2.3) and Gallati (2003: sect. 5.8).

[2] Besides risk reporting to senior management and shareholders, VaR is applied for allocating financial resources and risk-adjusted performance evaluation [cf. Jorion (1997:chap. 1)]. Furthermore, with the advent of the internal model approach [Basle Committee (1995, 1996a)], banks in the main financial jurisdictions may use their in-house VaR models for calculation of regulatory market-risk capital requirements.

[3] Bao, Lee, and Saltoglu only use existing models and do not propose any extensions. One consequence of this, for example, is their finding that the two CaViaR models they considered do not perform well under all circumstances—which precisely agrees with our results, based on different data. However, by a judicious choice of model extension, we demonstrate that CaViaR-based models can have attractive performance properties. A second consequence involves their finding that EVT models applied to Gaussian-GARCH filtered returns data only perform well at the lower quantiles—which, interestingly enough, again agrees with our results. By relaxing the Gaussian assumption, we are able to suggest a model that performs well across the entire quantile range between 1% and 5%.

in that article, none of the methods considered herein, namely generalized auto-regressive conditional heteroskedasticity (GARCH), mixed normal-GARCH (MN-GARCH), extreme value theory (EVT), and conditional autoregressive VaR (CAViaR), are used; and Brooks et al. (2005), who also use a variety of models (but none of which is identical to those in our study), including variants of the unconditional EVT approach and the regular GARCH model coupled with resampling strategies, and also a variety of nonparametric tail estimators.[4]

To the extent that a commercial bank and the regulator are interested in the aggregate VaR across different trading activities, the question arises whether first to aggregate profit and loss data and proceed with a univariate forecast model for the aggregate, or to start with disaggregate data. The latter approach leads to multivariate structural portfolio VaR models. In principle, these have the advan-tage of being suitable for sensitivity and scenario analysis and of conveying information about the structure of risk within a portfolio. However, as Berkowitz and O'Brien (2002) show in a recent study using actual commercial banks' VaR models and their profit and loss data, the aggregation and modeling problems involved may easily render a structural attempt to aggregate VaR poor for fore-casting purposes. While large banks and other financial institutions will ulti-mately require a reliable multivariate approach for some purposes, there are situations for which the univariate approach is adequate (e.g., certain growth funds, index-tracking funds, or when the focus is on forecasting aggregate VaR only). In their sample, for instance, Berkowitz and O'Brien (2002) illustrate that the complicated structural models were not able to outperform a simple univari-ate normal autoregressive moving average (ARMA)-GARCH model estimated on the aggregate profit and loss data when the aim was to forecast aggregate portfolio VaR. In addition, the reduced-form GARCH model turned out to yield less conservative VaR forecasts and hence would have been cheaper to imple-ment. Univariate models therefore are at least a useful complement to large structural models and may even be sufficient for forecasting portfolio VaR. Finally, establishing which univariate models appear most promising (and least so) could help in deciding which multivariate models are worth pursuing (and which not). Consequently we restrict attention to the univariate case.

For implementing univariate VaR-based measures, one seeks a precise quan-tile estimate relatively far out in the left tail of the return distribution for some specified future date. Existing approaches for obtaining such an estimate may be classified as follows: historical simulation simply utilizes empirical quantiles based on the available (possibly prefiltered) past data; fully parametric models describe the entire distribution of returns, including possible volatility dynamics; extreme value theory parametrically models only the tails of the return distribu-tion; and, finally, quantile regression directly models a specific quantile rather than the whole return distribution.

---

[4] Overall, the results in Brooks et al. (2005) favor the EVT and GARCH approaches, though their conclu-sions are based only on the 5% VaR using a holdout period of only 250 days, for three daily series from the London futures market over the period 1991 to 1997.

Below, we provide out-of-sample performance comparisons of models arising from these alternative strategies. The assessment is based on daily return data on the NASDAQ Composite Index, which is a typical representative of a portfolio of volatile financial assets. We show that, with only a few exceptions, all of the methods perform less than is desirable from a statistical viewpoint, although this may go undetected by the current regulatory assessment rules for model adequacy. In addition, we advance methodological concepts to VaR modeling by extending the EVT framework as applied in McNeil and Frey (2000) and by introducing a new specification for VaR quantile regressions. Both measures lead to significant improvements in out-of-sample VaR forecasts.

The remainder of the article is organized as follows. Section 1 briefly summarizes the major statistical approaches to VaR estimation. Section 2 examines methods for testing the adequacy of VaR forecasts. In Section 3 we describe the data and discuss the empirical results of the alternative forecasting methods. The final section provides concluding remarks and briefly mentions some of the issues relevant for multistep prediction. Some technical details are given in the appendix.

1   STATISTICAL APPROACHES TO VaR

In practice, VaR prediction is hampered by the fact that financial returns exhibit "nonstandard" statistical properties. Specifically, they are not independently and identically distributed (iid) and, moreover, they are not normally distributed. This is reflected by three widely reported stylized facts: (i) volatility clustering, indicated by high autocorrelation of absolute and squared returns; (ii) substantial kurtosis, that is, the density of the unconditional return distribution is more peaked around the center and possesses much fatter tails than the normal density; and (iii) mild skewness of the returns, possibly of a time-varying nature [see, e.g., Harvey and Siddique (1999), Rockinger and Jondeau (2002)]. As a consequence, "standard" methods, based on the assumption of iid-ness and normality, tend not to suffice, which has led to various alternative strategies for VaR prediction. The most prominent and/or most promising of these are outlined in the following subsections.

1.1   Historical Simulation

Arguably the simplest way to estimate VaR is to use the sample quantile estimate based on historic return data, which is referred to as historical simulation (HS). There are several varieties of this method, with various advantages and disadvantages [see Dowd (2002:sect. 4.5), Christoffersen (2003:chap. 5), and the references therein for detailed discussion]. We entertain the most popular way, which we call (naive) HS, and the most successful way, which is filtered historical simulation (FHS).

For HS, the VaR estimate for $t + 1$ is given by the empirical λ-quantile, $\hat{Q}_\lambda(\cdot)$, of a moving window of w observations up to date t, that is, $\widehat{VaR}_{t+1} = -\hat{Q}_\lambda(r_t, r_{t-1}..., r_{t-w+1})$. For example, with a moving window of length, say,

$w = 1000$ observations, the 5% VaR estimate is simply the negative of the 50th sample order statistic. Notice that, besides ignoring the oftentimes blatant non-iid nature of the data, predictions extending beyond the extreme returns observed during the past w observations are not possible with this method. Also, the resulting VaR estimates can exhibit predictable jumps when large negative returns either enter into or drop out of the window.

For FHS, a location-scale model such as Equations (3) and (4) below is used to prefilter the data. VaR forecasts are then generated by computing the VaR from paths simulated using draws from the filtered residuals. Barone-Adesi, Giannopoulos, and Vosper (1999, 2002) and Pritsker (2001) show that this method performs rather well, which agrees with our findings, as detailed below.

1.2 Fully Parametric: Location–Scale

Fully parametric models in the location-scale class are based on the assumption that returns belong to a location-scale family of probability distributions of the form

$$r_t = \mu_t + \epsilon_t = \mu_t + \sigma_t z_t, \tag{2}$$

where location μ_t and scale σ_t are $\mathcal{F}_{t-1}$-measurable parameters and $z_t \overset{\text{iid}}{\sim} f_Z(\cdot)$, where f_Z is a zero-location, unit-scale probability density that can have additional shape parameters (such as the degrees of freedom parameter in the Student's t distribution). The original ARCH and GARCH models took the z_t to be Gaussian, though this assumption was soon realized to be inadequate. Its replacement with a fat-tailed, possibly skewed distribution was a natural and quite effective extension. Many candidate distributions have been entertained; see, for example, the survey of Palm (1996) and the references therein for details.

The h-period-ahead VaR forecast based on information up to time t is $\widehat{\text{VaR}}_{t+h} = -(\hat{\mu}_{t+h} + \hat{\sigma}_{t+h} Q_\lambda(z))$, where $Q_\lambda(z)$ is the λ-quantile implied by f_Z. Approaches differ with respect to specification of the conditional location, μ_{t+h}, the conditional scale, σ_{t+h}, and the density, f_Z.

Unconditional parametric models set $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma$, thereby assuming that the returns are iid with density $\sigma^{-1} f_Z(\sigma^{-1}(r_t - \mu))$. Conditionally homoskedastic parametric models allow for a time-varying conditional mean, possibly captured by an ARMA(p, q) process, that is,

$$\mu_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i} + \sum_{j=1}^{q} b_j \, \epsilon_{t-j}, \tag{3}$$

with $\sigma_t \equiv \sigma$, $t = 1, \ldots, T$. In light of the observed volatility clustering, this model class will be only of marginal use. Instead, conditionally heteroskedastic parametric models, which allow the scale parameter to be a function of past information, are frequently used. The most popular formulation is the GARCH(r, s) model

$$\sigma_t^2 = c_0 + \sum_{i=1}^{r} c_i \, \epsilon_{t-i}^2 + \sum_{j=1}^{s} d_j \sigma_{t-j}^2, \tag{4}$$

introduced by Bollerslev (1986).

In the empirical analysis below, we utilize three different assumptions for the innovation distribution, f_Z, in Equation (2): the normal; the Student's t with $\nu \in \mathbb{R}_+$ degrees of freedom (in short, t distribution); and the generalized asymmetric t (in short, skewed t), with density

$$f(z; d, \nu, \theta) = C \left(1 + \frac{(-z\theta)^d}{\nu} \right)^{-\frac{\nu+1}{d}} \mathcal{I}(z < 0) + C \left(1 + \frac{(z/\theta)^d}{\nu} \right)^{-\frac{\nu+1}{d}} \mathcal{I}(z \geq 0), \tag{5}$$

where $d, v, \theta \in \mathbb{R}_+$, $\mathcal{I}(\cdot)$ is the indicator function, $C = [(\theta + \theta^{-1})d^{-1}\nu^{1/d}B(d^{-1}, \nu)]^{-1}$, and $B(\cdot, \cdot)$ denotes the beta function. The rth raw integer moment, $0 \leq r < \nu d$, for the skewed t is

$$\frac{(-1)^r \theta^{-(r+1)} + \theta^{r+1}}{\theta^{-1} + \theta} \frac{B\left(\frac{r+1}{d}, \nu - \frac{r}{d}\right)}{B\left(\frac{1}{d}, \nu\right)} \nu^{r/d},$$

from which, for example, variance, skewness, and kurtosis can be computed if they exist. The cumulative distribution function (cdf) of the skewed t (as required for VaR calculation) is given by

$$F(z) = \begin{cases} \frac{I_L(\nu, 1/d)}{1 + \theta^2}, & \text{if } z \leq 0, \\ \frac{I_U(1/d, \nu)}{1 + \theta^{-2}} + (1 + \theta^2)^{-1}, & \text{if } z > 0, \end{cases}$$

where $L = \nu / \left[\nu + (-z\theta)^d \right]$, $U = (z/\theta)^d / \left[\nu + (z/\theta)^d \right]$, and

$$I_x(a,b) = \frac{B_x(a,b)}{B(a,b)} = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt \quad (a, b > 0)$$

is the incomplete beta ratio. GARCH-type models coupled with the skewed t distribution have been frequently found to-deliver excellent forecast results; see, for example, Mittnik and Paolella (2000), Giot and Laurent (2004), and the references therein.

1.3 Fully Parametric: Dynamic Feedback

A less obvious parametric alternative to the location-scale model discussed above is to link a GARCH-type structure to a discrete mixture of normal distributions, allowing for dynamic feedback between the normal components. Numerous authors have empirically shown that a mixture of normals (with between two and four components) can fit the unconditional distribution of asset returns extremely well, and, more recently, several ways of using the mixed normal distributional assumption with GARCH-type structures have been considered. These are reviewed in Haas, Mittnik, and Paolella (2004a, b), in which a general

model structure is proposed that nests all of these attempts, and general stationarity conditions are derived. With one component, the model reduces to the usual GARCH with normal innovations. The model is appealing because it allows for an economic interpretation in terms of information flows between groups of agents [see the discussion and references in Haas, Mittnik, and Paolella (2004a)] and, from a practical viewpoint, was shown to deliver competitive VaR forecasts.

Briefly, time series $\{\epsilon_t\}$ is generated by an n-component mixed normal GARCH(r, s) process (denoted MixN-GARCH) if the conditional distribution of ϵ_t is an n-component mixed normal with zero mean, that is,

$$\epsilon_t | \mathcal{F}_{t-1} \sim \text{MN}\left(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right), \tag{6}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$, $\boldsymbol{\sigma}_t^2 = (\sigma_{1t}^2, \ldots, \sigma_{nt}^2)$, and the mixed normal density is given by

$$f_{\text{MN}}\left(y; \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma}_t^2\right) = \sum_{j=1}^{n} \omega_j \phi\left(y; \mu_j, \sigma_{jt}^2\right),$$

ϕ is the normal pdf, $\omega_j \in (0, 1)$ with $\sum_{j=1}^{n} \omega_j = 1$ and, to ensure zero mean, $\mu_n = -\sum_{j=1}^{n-1}\left(\omega_j / \omega_n\right)\mu_j$. The "law of motion" for the component variances, denoted by $\sigma_t^{(2)}$, is given by the GARCH-like structure

$$\boldsymbol{\sigma}_t^{(2)} = \gamma_0 + \sum_{i=1}^{r} \boldsymbol{\gamma}_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \boldsymbol{\Psi}_j \boldsymbol{\sigma}_{t-j}^{(2)}, \tag{7}$$

where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{in})'$, $i = 0, \ldots, r$, are $n \times 1$ vectors, and $\boldsymbol{\Psi}_j$, $j = 1, \ldots, s$, are $n \times n$ matrices. We restrict $\boldsymbol{\Psi}_j$ to be diagonal, which, as discussed in Haas, Mittnik, and Paolella (2004a), yields a much more parsimonious model with little loss in the quality of in-sample fit and forecasting ability.[5]

It is plausible that the component of the mixture assigned to the most volatile observations does not require a GARCH structure, that is, occasionally occurring jumps in the level of volatility may be captured by a component with a relatively large, but constant, variance. We denote by MixN(n, g) the model given by Equations (6) and (7), with n component densities, but such that only g, $g \leq n$, follow a GARCH(1,1) process (and $n - g$ components are restricted to be constant).

In the empirical work in Section 4, we take $r = s = 1$, and, regarding n and g, consider the three cases MixN(2, 2), MixN(3, 2), and MixN(3, 3). As for the location-scale models in Section 2.2, an ARMA structure can also be imposed onto the returns; in the AR(1) case, as used below, this is $r_t = a_0 + a_1 r_{t-1} + \epsilon_t$.

---

[5] For $n > 1$, the MN-GARCH model (with or without the diagonal restriction on the $\boldsymbol{\Psi}_j$ matrices) has the interesting property that the skewness implied by the (fitted) model is time varying. Thus the MN-GARCH model can also account for this stylized fact, and does so in a different way than otherwise considered in the literature.

Anticipating our empirical results, the MixN(3, 3) and, particularly for smaller sample sizes, the MixN(3, 2), perform reasonably well, though they are still inferior to some of the other models entertained. Similar to replacing the normal assumption in a standard GARCH model, one ad hoc remedy is to replace the mixed normality assumption by a more flexible (symmetric) distribution. For example, Haas, Mittnik, and Paolella (2004a) used a Student's t for each component. In this article, we consider instead the generalized exponential distribution (GED) for which the normal is a genuine special case, instead of a limiting case. In addition to fatter tails, it also allows for thinner tails, which we will see below may be beneficial in some circumstances.

The location-zero, scale-one GED density with exponent p is given by

$$f(x;p) = \frac{p}{2\Gamma(p^{-1})}\exp\{-|x|^p\}, \quad p \in \mathbb{R}_+. \tag{8}$$

For $p > 2$, the tails are thinner than those of the normal. After rescaling, the normal and Laplace distributions arise as special cases for $p = 2$ and $p = 1$, respectively. As $p \to \infty$, the GED approaches a uniform distribution. The cdf is required for VaR calculations; straightforward calculation shows that the cdf for $x \le 0$ is given by

$$F(x;p) = \frac{1}{2}\left(1 - \bar{\Gamma}_{(-x)^p}\left(p^{-1}\right)\right), \quad x \le 0, \tag{9}$$

where $\bar{\Gamma}$ is the incomplete gamma ratio. The symmetry of the density implies that $F(x) = 1 - F(-x)$, from which $F(x)$ for $x > 0$ can be computed using Equation (9).

We denote by MixGED(n, g) the described GARCH(1,1) mixture with n GED components, each with shape parameter p_i, $i = 1, \ldots, n$. These additional n parameters are (as usual) not prespecified, but jointly estimated along with the remaining ones.

1.4   Extreme Value Theory

Extreme value theory is concerned with the distribution of the smallest- and largest-order statistics and focuses only on the tails of the return distribution. A comprehensive overview of the subject is provided by Embrechts, Klüppelberg, and Mikosch (1997), while Christoffersen (2003) gives a highly accessible and streamlined account. We follow convention and restrict attention to the right tail, an implication of which is that, if the left tail of the data is of interest (as is more often the case in a financial risk context), then the EVT analysis should be applied to the absolute value of the negative returns.

To briefly review the concepts that will be required in the subsequent analysis, let $\{X_t\}_{t=1}^T$ be a sequence of iid random variables, and $M_T = \max(X_1, X_2, \ldots, X_T)$. If there exist norming constants $c_T > 0$ and $d_T \in \mathbb{R}$ and some nondegenerate distribution function H such that

$$\frac{M_T - d_T}{c_T} \xrightarrow{d} H, \tag{10}$$

then, for $1 + \xi x > 0$,

$$H_\xi(x) = \begin{cases} \exp\left\{-(1 + \xi x)^{-\frac{1}{\xi}}\right\}, & \text{if} \quad \xi \neq 0, \\ \exp\{-\exp\{-x\}\}, & \text{if} \quad \xi = 0, \end{cases}$$

where H_ξ is called the generalized extreme value (GEV) distribution [cf. Embrechts, Klüppelberg, and Mikosch (1997: 121, 152)]. In other words, there is one, and only one, possible family of limiting distributions for sample maxima. The random variable X, with distribution function F, is then said to belong to the maximum domain of attraction of an extreme value distribution; in short, $F \in \text{MDA}(H_\xi)$.

Parameter ξ is crucial because it governs the tail behavior of $F(x)$. Distributions $F \in \text{MDA}(H_\xi)$ are heavy tailed for $\xi > 0$, which includes inter alia the Pareto distribution and the stable Paretian distributions. For $\xi = 0$, the tail decreases at an exponential rate, as is the case for the normal distribution, while distributions with $\xi < 0$ have a finite right end point. Indeed, $\text{MDA}(H_\xi)$ includes essentially all the common continuous distributions occurring in applied statistics.

Consider now the distribution function of excesses, $Y = X - u$, of the iid random variable X over a high, fixed threshold u, that is,

$$F_u(y) = \text{P}(X - u \leq y \mid X > u), \quad y \geq 0.$$

For excesses over thresholds, a key result, due to Pickands (1975), is that the generalized Pareto distribution (in short, GPD)

$$G_{\xi,\beta}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\beta}\right)^{-\frac{1}{\xi}}, & \text{if} \quad \xi \neq 0, \\ 1 - e^{-\frac{y}{\beta}}, & \text{if} \quad \xi \neq 0, \end{cases} \tag{11}$$

with support

$$y \geq 0, \qquad \text{if} \quad \xi \geq 0,$$
$$0 \leq y \leq -\beta/\xi, \quad \text{if} \quad \xi < 0,$$

and scale parameter β, arises naturally as the limit distribution of scaled excesses of iid random variables over high thresholds.

Two main strands of the current literature exist. The first assumes fat-tailed data and makes use of an estimator for the tail index [see, e.g., Danielsson and de Vries (2000)]. Below, we focus on the second and more general strand of literature, which makes use of the limit result for peaks over thresholds (POT) in Equation (11) and is not confined to fat-tailed data. Suppose that the X_t are iid with distribution function $F \in \text{MDA}(H_\xi)$. Then, for a chosen threshold $u = X_{k+1,T}$ given by the $(k + 1)$st descending order statistic, define

$$F_u(y) = \text{P}(X - u \leq y \mid X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad y \geq 0,$$

which can be rewritten as

$$\bar{F}(u + y) = \bar{F}(u)\bar{F}_u(y). \tag{12}$$

In Equation (12), $\bar{F}(u)$ can be estimated by its empirical counterpart, $\bar{F}_T(u) = k/T$, with $F_T(u)$ being the empirical distribution function of X. For a high enough threshold,

$$\bar{F}_u(y) \approx 1 - G_{\xi,\beta(u)}(y), \tag{13}$$

so that, given estimates $\hat{\xi}$ and $\hat{\beta}$, $1 - G_{\hat{\xi},\hat{\beta}}(y)$ provides an estimate for $\bar{F}_u(y)$. Thus the tail probability for $X > u$ can be estimated by

$$\widehat{\bar{F}(x)} = \frac{k}{T}\left(1 + \hat{\xi}\,\frac{x - u}{\hat{\beta}}\right)^{-1/\hat{\xi}}. \tag{14}$$

A quantile estimator, $F(x_p) > 1 - k/T$, is obtained by inverting Equation (14), that is,

$$\hat{x}_{p,k} = X_{k+1,T} + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-p}{k/T}\right)^{-\hat{\xi}} - 1\right), \tag{15}$$

recalling $u = X_{k+1,T}$. Here the choice of k suffers from similar problems as for the Hill estimator. Choosing u too high leads to very few exceedances, and thus a high variance for the estimator, while low threshold values induce bias, as Equation (13) works well only in the tail.[6]

    Although EVT is a natural candidate for VaR modeling, in light of the aforementioned stylized facts, EVT's iid assumption is inappropriate for most asset return data. While extensions that relax the independence assumption exist [cf. Embrechts, Klüppelberg, and Mikosch (1997: 209ff)], one may alternatively apply the EVT analysis to appropriately filtered data. Diebold, Schuermann, and Stroughair (1998) propose fitting a time-varying volatility model to the data and then estimating the tail of the filtered or standardized residuals, $z_t = (r_t - \mu_t)/\sigma_t$, by an EVT model. This yields an estimate for the standardized quantile, $Q_\lambda(z)$, as defined by Equation (1), and thus for the VaR,

$$\mathrm{VaR}_t = -(\mu_t + \sigma_t Q_\lambda(z)). \tag{16}$$

With a correct model specification of the location and scale dynamics and use of consistent parameter estimates, the filtered model residuals will be approximately iid, as assumed in EVT modeling.

    The Gaussian AR(1)-GARCH(1,1) filter, as applied by McNeil and Frey (2000) in this context, is just a special case that, while capable of removing the majority of volatility clustering and rendering the data approximately iid, will almost always be a misspecified model for financial return data of daily or higher frequency.

---

[6] However, simulation evidence in McNeil and Frey (2000) for Student's t data suggests that the mean-squared error for quantile estimates based on POT is far less sensitive to the choice of the threshold than for the Hill estimator.

Much of the misspecification can be accommodated for by using a fat-tailed and asymmetric distribution such as the skewed t for f_z, given in Equation (5). As such, its use would be expected to result in more accurate AR and GARCH parameter estimates, as well as filtered (estimated) σ_t values, which in turn lead to improved scale forecasts, $\hat{\sigma}_{t+h}$.

On the other hand, Bollerslev and Wooldridge (1992) show that if the conditional mean and volatility dynamics are properly specified, then the conditional mean and volatility are consistently estimated by pseudo-maximum likelihood—that is, maximum-likelihood estimation under normality assumptions, even when innovations are not normally distributed.[7] Because the proper specification of the volatility dynamics is clearly an unattainable goal for actual return series, it is far from obvious which specification will be optimal, and rather, the decision should be based on out-of-sample VaR forecasting performance.

1.5 Quantile Regression Approach

The determination of VaR naturally lends itself to the concept of quantile regression. To estimate conditional quantiles, the time series of the specified quantile is explicitly modeled using any information deemed to be relevant. No distributional assumptions for the time series behavior of returns is needed. The basic idea is to model the conditional λ-quantile, $Q_\lambda(r_t \mid X_t) = -\text{VaR}_t$, as some function of the information $X_t \in \mathcal{F}_{t-1}$, that is,

$$\text{VaR}_t \equiv -g_\lambda(\mathbf{X}_t; \boldsymbol{\beta}_\lambda), \tag{17}$$

where $g(\cdot, \cdot)$ and parameter vector $\boldsymbol{\beta}$ explicitly depend on λ. A good choice of relevant information and of the functional form should yield a close approximation to the population quantile [cf. Chernozhukov and Umantsev (2001)]. Koenker and Basset (1978) generalize the common linear regression framework by shifting the focus from the conditional mean to conditional quantiles. As shown, for example, in Koenker and Portnoy (1997), the unconditional sample λ-quantile, $\lambda \in (0, 1)$, can be found as the solution to

$$\min_{\beta \in \mathbb{R}} \left\{ \sum_{r_t \geq \beta} \lambda |r_t - \beta| + \sum_{r_t < \beta} (1 - \lambda)|r_t - \beta| \right\}. \tag{18}$$

Extending this to the classical linear regression freamwork, Koenker and Bassett (1978) define the λth regression quantile estimator by

$$\hat{\boldsymbol{\beta}}(\lambda) = \underset{\boldsymbol{\beta} \in \mathbb{R}^k}{\text{argmin}} \left\{ \sum_{r_t \geq \mathbf{x}_t' \boldsymbol{\beta}} \lambda |r_t - \mathbf{x}_t' \boldsymbol{\beta}| + \sum_{r_t < \mathbf{x}_t' \boldsymbol{\beta}} (1 - \lambda)|r_t - \mathbf{x}_t' \boldsymbol{\beta}| \right\},$$

---

[7] Their proofs rest on the existence of conditional variances. This seemingly innocuous assumption would, for example, rule out use of the (skewed) t with degrees of freedom less than two and also the asymmetric stable Paretian distribution.

where the x_t are nonrandom vectors.[8] The key assumption in the linear quantile regression model is that $r_t = x'_t \beta_\lambda + u_{t,\lambda}$. Note that the distribution of the error term is left unspecified. The only assumption made is that the conditional quantile function is given by $Q_\lambda(r_t | x_t) = x'_t \beta_\lambda$, and thus $Q_\lambda(u_{t,\lambda} | x_t) = 0$.[9]

One natural extension of the objective function for the general, possibly non-linear case of Equation (17), proposed by Engle and Manganelli (2004), is

$$\min_{\beta \in \mathbb{R}^k} \left\{ \sum_{r_t \geq \mathrm{VaR}_t} \lambda |r_t + \mathrm{VaR}_t| + \sum_{r_t < -\mathrm{VaR}_t} (1 - \lambda)|r_t + \mathrm{VaR}_t| \right\} \qquad (19)$$

with, according to Equation (17), $\mathrm{VaR}_t \equiv -g(x_t; \beta_\lambda)$ or, in the linear case, $\mathrm{VaR}_t \equiv x'_t \beta_\lambda$. Consistency and asymptotic normality of the nonlinear regression quantiles for the time-series case are established in Engle and Manganelli (2004).

Chernozhukov and Umantsev (2001) use quantile regressions to model VaR—without, however, examining the model performance in terms of the sequence of VaR violations, as is done below. Taylor (1999) deals with the estimation of multiperiod VaR in the context of exchange rates, specifying Equation (17) as linear functions of (transforms of) volatility estimates and the return horizon. As is common in the VaR literature, Taylor (1999) judges the efficiency of VaR estimates only on the basis of unconditional coverage (to be defined in Section 2 below).

Because our focus is exclusively on one-step forecasting performance, we more closely examine the conditional VaR approach formulated in Engle and Manganelli (2004), which is amenable to our maintained assumption that return data contain sufficient information for forecasting. In their specification of Equation (17), they link VaR to the conditional standard deviation of the returns such that an increase in the latter leads to a more dispersed return distribution and thus, ceteris paribus, to a higher VaR. Their CAViaR specifications include VaR_{t-1} as an explanatory variable in x_t, to adapt to serial dependence in volatility and mean. A function of r_{t-1} is also included to link the conditional quantile to return innovations.

As mentioned above, no explicit distributional assumptions need to be made, guarding against this source of model misspecification. Although many specifications of Equation (17) are conceivable, we first adopt those put forth in Engle and Manganelli (2004). The baseline CAViaR model is given by

---

[8] Their proof of consistency and joint asymptotic normality relies on iid error terms with a continuous distribution function and fixed regressors. Nonlinear absolute ARCH-power specifications are the subject of Koenker and Zhao (1996). For more general asymptotic results, see Chernozhukov and Umantsev (2001) and the references therein.

[9] The well-known least absolute deviation (LAD) estimator for the regression median arises as the special case $\lambda = 0.5$. It yields more efficient estimates for the population mean than least squares in the presence of fat-tailed error distributions [see Bassett and Koenker (1978) for the iid error case with fixed regressors].

$$\text{VaR}_t = \text{VaR}_{t-1} + \beta[\mathcal{I}(r_{t-1} \leq -\text{VaR}_{t-1}) - \lambda].$$

As typically $\lambda \leq 0.05$ for risk management purposes, we have an asymmetric response: VaR_t will jump upward when a violation occurs and will slowly decrease otherwise—provided that the a priori conjecture $\beta > 0$ holds. In the baseline model, the adaptive process "learns" nothing from the actual size of returns (except whether or not the returns are in line with VaR), as is the case with the symmetric absolute value CAViaR specification,

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 |r_{t-1}|.$$

It allows the autoregressive parameter, β_1, to be different from one, and introduces a direct response of the quantile to the return process, treating the effect of extreme returns on VaR—and implicitly, on volatility—symmetrically. The symmetric assumption is relaxed in the asymmetric slope CAViaR specification,

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \beta_2 \max[r_{t-1}, 0] + \beta_3 \max[-r_{t-1}, 0],$$

which allows the VaR prediction to respond asymmetrically to positive and negative returns, and so can accommodate the leverage effect. The indirect GARCH(1,1) CAViaR process,

$$\text{VaR}_t = \left(\beta_0 + \beta_1 \text{VaR}_{t-1}^2 + \beta_2 r_{t-1}^2\right)^{1/2}, \tag{20}$$

proposed by Engle and Manganelli (2004), would be appropriate if the data were generated by a location-scale model [Equation (2)], with a GARCH(1, 1) process for the conditional scale, σ_t, and with zero location parameter, μ_t.

Autocorrelation in financial returns is often nonnegligible [see, e.g., Danielsson and Morimoto (2000)]. This property can be incorporated by extending the existing CAViaR framework by allowing the returns to have a time-varying mean of the form

$$\mu_t = \text{E}(r_t \mid \mathcal{F}_{t-1}), \tag{21}$$

which may, for example, be captured by a regression, ARMA, or ARMAX model. An indirect GARCH specification of orders r and s with conditional mean μ_t can then be written as

$$\left(\frac{\text{VaR}_t + \mu_t}{z_\lambda}\right)^2 = c_0 + \sum_{i=1}^{r} c_i (r_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{s} d_j \left(\frac{\text{VaR}_{t-j} + \mu_{t-j}}{z_\lambda}\right)^2. \tag{22}$$

The indirect conditional mean GARCH(r, s) CAViaR model of Equations (21) and (22) reduces to that in Engle and Manganelli (2004) for $r = s = 1$ and $\mu_t = 0$. In the application below, it will be demonstrated that this more general CAViaR specification leads to a significant improvement in performance. There, we specify an AR(1)

model, $r_t = ar_{t-1} + \epsilon_t$ for Equation (21), and GARCH orders $r = s = 1$, leading to the indirect AR(1)-GARCH(1,1) CAViaR,[10]

$$\text{VaR}_t = -ar_{t-1} + \left(\beta_0 + \beta_1 (\text{VaR}_{t-1} + ar_{t-2})^2 + \beta_2 (r_{t-1} - ar_{t-2})^2 \right)^{1/2}, \quad \beta_i > 0. \ (23)$$

1.6 Other Approaches

With respect to the paradigm outlined in Section 1.2, there have been many proposed GARCH formulations and numerous distributional assumptions explored in the literature. Good overviews are provided by Bao, Lee, and Saltoglu (2003), who examine and compare the forecasting ability of 13 distributional assumptions crossed with nine volatility structures, and Hansen and Lunde (2004), who consider 330 such models. We also made use of such "combinations," and summarize the results now. First, we entertained the popular assymetric power ARCH (APARCH) model proposed by Ding, Granger, and Engle (1993), which nests several well-known GARCH structures including Equation (4), and allows for a leverage effect; and second, we entertained the use of the asymmetric stable Paretian distributional assumption (incorporated into both the GARCH and APARCH models). With respect to the dataset analyzed in Section 3, the more general APARCH structure actually led to a decrease in forecast quality compared to Equation (4), as did the absolute value GARCH model advocated by Nelson and Foster (1994) (which is also subsumed in the APARCH model). The stable Paretian distributional assumption, while better than both the normal and the t [agreeing with the findings of Mittnik and Paolella (2003)], did not outperform the use of the skewed t, the latter of which is more straightforward to program and much faster to estimate.

There are other model classes suitable for VaR prediction that we did not consider. The major ones include:

1. Long memory/fractionally integrated GARCH-type (FIGARCH) models. A summary of the major models and references to the literature are provided by the valuable review article of Poon and Granger (2003). Regarding their forecasting ability, there appear to be mixed results. In their large study, Bao, Lee, and Baltoglu (2003) report that the distributional assumption for the innovations of a GARCH model is

---

[10] To see this, when the process generating the returns is a GARCH(1, 1) with the AR(1) mean equation $r_t = ar_{t-1} + \varepsilon_t$ with $a_0 \equiv 0$. Then $\sigma_t^2 = c_0 + c_1 (r_{t-1} - \mu_{t-1})^2 + d_1 \sigma_{t-1}^2$, where $c_0, c_1, d_1 > 0$. Substitution of $\text{VaR}_{t-1} = -\mu_{t-1} - \sigma_{t-1} z_\lambda$ yields

$$\sigma_t^2 = c_0 + c_1 (r_{t-1} - \mu_{t-1})^2 + d_1 S_{t-1}, \quad S_{t-1} = \left(\frac{\text{VaR}_{t-1} + \mu_{t-1}}{z_\lambda} \right)^2,$$

and with $\mu_t = a_1 r_{t-1}$ and $\text{VaR}_t = -ar_{t-1} - z_\lambda \left(c_0 + c_1 (r_{t-1} - ar_{t-2})^2 + d_1 S_{t-1} \right)^{1/2}$. Taking z_λ (which is a constant for iid innovations) into the root and noting that $z_\lambda < 0$ for small λ (so $z_\lambda = -\sqrt{z_\lambda^2}$), we obtain the desired CAViaR expression of Equation (23) after appropriately relabelling the parameters.

far more influential than the GARCH specification, including FIGARCH. Vilasuso (2002) demonstrates superior forecast performance from a FIGARCH model for exchange rates compared to GARCH and IGARCH. However, he allows only for normal innovations, which severely diminishes the value of such findings. In their large forecasting study, Lux and Kaizoji (2004) document "occasional dramatic failures" of FIGARCH, but report good performance of multifractal processes, which also exhibit long memory (see below).

2. Markov-switching GARCH models. A recent overview and discussion of the many contributions in this category are given in Haas, Mittnik, and Paolella (2004b). There, a Markov-switching GARCH model is introduced that is straightforward to estimate and admits tractable stationarity conditions. In a forecasting comparison, they find that the new model performs about as well as the MixN-GARCH model, which can be seen as a special case of this, though the latter was generally favored.

3. Use of realized volatility. Based on intraday data, daily realized volatility can be "observed" (i.e., independent of a model and essentially error free) and then used for daily prediction purposes [see Martens (2001), Giot and Laurent (2004), Galbraith and Kisinbay (2005), Koopman, Jungbacker, and Hol (2005), and the references therein]. Giot and Laurent (2004) demonstrate with a variety of datasets that the method does not lead to improvements in forecast quality when compared to use of a skewed-t A-PARCH model for daily returns. The comparison studies of Galbraith and Kisinbay (2005) and Koopman, Jungbacker, and Hol (2005) concentrate on volatility and variance prediction instead of VaR, and, from their respective choice of models, both conclude that realized volatility is the best choice.[11]

4. Markov-switching multifractal processes. This promising class of models, with its origins in fluid dynamics by Benoît Mandelbrot in the 1970s, appears particularly well suited for medium- and long-term forecasting [see Calvet and Fisher (2004), Lux (2004), and Lux and Kaizoji (2004), for details and further references].

Three more categories are worth mentioning via their relation to volatility prediction, though much less is known about their use for VaR forecasting:

---

[11] As in all comparison studies, the choice of models will never be complete. For example, with respect to GARCH-type models, in both the aforementioned articles, the authors use only the plain, normal GARCH(1,1) model. In this article and numerous previous studies, the normal GARCH model has been shown to be vastly inferior to simple improvements, such as the skewed-t GARCH or skewed-t APARCH.

5. Implied volatility models. A detailed account of volatility prediction based on option prices is given in Poon and Granger (2003). Based on the articles they review, there is favorable evidence that this model class produces competitive volatility forecasts. See also Koopman, Jungbacker, and Hol (2005) for comparisons with other model classes.

6. Stochastic volatility models. Some overview of this field is provided by Poon and Granger (2003) and Durbin and Koopman (2001: sect. 10.6), while the forthcoming collection by Shephard (2005) should prove most valuable.

7. Multivariate GARCH (MGARCH). It is natural to expect that modeling the comovement of financial time series would enhance predictive ability. However, evidence against this is provided by Brooks and Persand (2003), which appears to be the first comparison of various univariate models and a MGARCH for VaR and volatility forecasting. They conclude that the gain in using MGARCH models is minimal and is not worth the trouble because of their complexity and estimation difficulties. This complements the results of Berkowitz and O'Brien (2002), as discussed in the introduction.[12]

2 COMPARING AND TESTING THE FIT OF VaR MODELS

To assess the predictive performance of the models under consideration, we follow Christoffersen's (1998) framework, which is designed for evaluating the accuracy of out-of-sample interval forecasts. Defining $H_t = \mathcal{I}(r_t < -\text{VaR}_t)$, Christoffersen (1998) terms the sequence of VaR forecasts efficient with respect to $\mathcal{F}_{t-1}$ if

$$\text{E}[H_t \,|\, \mathcal{F}_{t-1}] = \lambda, \tag{24}$$

which, by applying iterated expectations, implies that H_t is uncorrelated with any function of a variable in the information set available at $t - 1$. If Equation (24) holds, then VaR violations will occur with the correct conditional and unconditional probability, and neither the forecast for VaR_t nor that for H_t could be improved.

Although a general test of Equation (24) is desirable, we follow Christoffersen (1998) in using intermediate statistics for testing specific implications of the general hypothesis, so that particular inadequacies of a model can be revealed. By specifying $\mathcal{F}_{t-1}$ to include at least $\{H_1, H_2, \ldots, H_{t-1}\}$, it is straightforward to show [Christoffersen (1998: Lemma 1)] that efficiency implies

$$H_t|\mathcal{F}_{t-1} \overset{\text{iid}}{\sim} \text{Ber}(\lambda), \quad t = 1, 2, \ldots, T, \tag{25}$$

---

[12] Because Brooks and Persand (2003) used a normality assumption for VaR prediction and only considered one MGARCH model [the diagonal VEC form from Bollerslev, Engle, and Wooldridge (1988)], their conclusions need to be tempered.

where Ber($\cdot$) denotes the Bernoulli distribution. Below, Equation (25) is referred to as correct conditional coverage.

2.1   Test of Unconditional Coverage

By taking iterated expectations, Equation (24) implies correct unconditional coverage of the interval forecasts. We test for the correct number of violations by

$$H_0 : E[H_t] = \lambda \quad \text{versus} \quad H_A : E[H_t] \neq \lambda.$$

Under the null hypothesis, Equation (25) implies the likelihood-ratio test statistic

$$\text{LR}_{\text{uc}} = 2\left[\mathcal{L}\left(\hat{\lambda}; H_1, H_2, \ldots, H_T\right) - \mathcal{L}(\lambda; H_1, H_2, \ldots, H_T)\right] \stackrel{\text{asy}}{\sim} \chi_1^2, \tag{26}$$

where $\mathcal{L}(\cdot) = \log L$ denotes the log binomial likelihood. The maximum-likelihood estimation $\hat{\lambda}$, is the ratio of the number of violations, n_1, to the total number of observations, $n_0 + n_1 = T$, that is, $\lambda = n_1/(n_0 + n_1)$.

2.2   Test of Independence

Value-at-risk forecasts that do not take temporal volatility dependence into account may well be correct on average, but will produce violation clusters [cf. Christoffersen (1998)], a phenomenon that is ignored when considering unconditional coverage.

Several tests for independence have been proposed in the literature, including the runs tests and the Ljung-Box test [Ljung and Box (1978)]. More recently, a test based on the time between exceedances was proposed in Danielsson and Morimoto (2000). Under the null hypothesis, a violation today has no influence on the probability of a violation tomorrow. Christoffersen (1998) models $\{H_t\}$ as a binary first-order Markov chain with transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \quad \pi_{ij} = P(H_t = j \mid H_{t-1} = i),$$

as the alternative hypothesis of dependence.[13] The approximate joint likelihood, conditional on the first observation, is

$$L(\Pi; H_2, H_3, \ldots, H_T \mid H_1) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}, \tag{27}$$

where n_{ij} represents the number of transitions from state i to state j, that is,

$$n_{ij} = \sum_{t=2}^{T} \mathcal{I}(H_t = i \mid H_{t-1} = j),$$

---

[13] The runs test is uniformly most powerful against this alternative [see, e.g., Lehmann (1986)]. We opt for the framework used here because it can be easily integrated into a test of the more general hypothesis of Equation (25).

and the maximum-likelihood estimators under the alternative hypothesis are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad \text{and} \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Under the null hypothesis of independence, we have $\pi_{01} = \pi_{11} \equiv \pi_0$, from which the conditional binomial joint likelihood follows as

$$L(\pi_0; H_2, \ldots, H_T \mid H_1) = (1 - \pi_{01})^{n_{00} + n_{10}} \pi_{01}^{n_{01} + n_{11}}. \tag{28}$$

The maximum-likelihood estimate, $\hat{\pi}_0$, is analogous to that in the unconditional coverage test, and the likelihood ratio (LR) test is given by

$$\text{LR}_{\text{ind}} = 2 \big[\mathcal{L} \big(\hat{\Pi}; H_2, \ldots, H_T \mid H_1 \big) - \mathcal{L}(\hat{\pi}_0; H_2, \ldots, H_T \mid H_1) \big] \overset{\text{asy}}{\sim} \chi_1^2. \tag{29}$$

2.3 Conditional Coverage

Because $\hat{\pi}_0$ is unconstrained, the test in Equation (29) does not take correct coverage into account. Christoffersen (1998) suggests combining Equations (26) and (29) to

$$\text{LR}_{\text{cc}} = 2 \big[\mathcal{L} \big(\hat{\Pi}; H_2, \ldots, H_T \mid H_1 \big) - \mathcal{L}(\lambda; H_2, \ldots, H_T \mid H_1) \big] \overset{\text{asy}}{\sim} \chi_2^2 \tag{30}$$

in order to test correct conditional coverage [Equation (25)]. By conditioning on the first observation in Equation (26), we have

$$\text{LR}_{\text{cc}} = \text{LR}_{\text{uc}} + \text{LR}_{\text{ind}},$$

which provides a means to check in which regard the violation series $\{H_t\}$ fails the correct conditional coverage property of Equation (25).[14]

2.4 Dynamic Quantile Test

Equation (24) is stronger than correct conditional coverage; it suggests that any $x_{t-1} \in F_{t-1}$ be uncorrelated with H_t. In particular, Engle and Manganelli (2004) remark that conditioning violations on the VaR for the period itself is essential. To illustrate this point, they let $\{\text{VaR}_t\}_{t=1}^T$ be a sequence of iid random variables such that

$$\text{VaR}_t = \begin{cases} K, & \text{with probability } 1 - \lambda, \\ -K, & \text{with probability } \lambda. \end{cases}$$

---

[14] As with all asymptotically motivated inferential procedures, the actual size of the tests for finite samples can deviate from their nominal sizes. Lopez (1997) examines the size of unconditional and conditional coverage tests via simulation, as well as their power against various model misspecifications. For a sample size of 500 observations, he finds both tests to be adequately sized. Even for such a small sample, power appears to be reasonable. For the LR_{uc} test, for example, he reports that, for λ values of 0.05 or smaller, if the true data generating process is conditionally heteroskedastic, then power is well above 60% for wrong distributional assumptions for the innovations. In general, the tests have only moderate power when volatility dynamics are closely matched but power increases under incorrect innovation distributions, especially further out in the tails.

Then, for K very large and conditioning also on VaR_t, the violation sequence exhibits correct conditional coverage, as tested by Equation (30), but, conditional on VaR_t, the probability of a violation is either almost zero or almost one. None of the above tests has power against this form of inefficiency.

To operationalize Equation (24), one can, similar to Christoffersen (1998) and Engle and Manganelli (2004), regress H_t on a judicious choice of explanatory variables in $\mathcal{F}_{t-1}$, for example,

$$H_t = \lambda_0 + \sum_{i=1}^{p} \beta_i H_{t-i} + \beta_{p+1} \widehat{\text{VaR}}_t + \mu_t,$$

where, under the null hypothesis, $\lambda_0 = \lambda$ and $\beta_i = 0$, $i = 1, \ldots, p+1$. In vector notation, we have

$$\mathbf{H} - \lambda\iota = \mathbf{X}\boldsymbol{\beta} + \mathrm{u}, \quad u_t = \begin{cases} -\lambda, & \text{with probability } 1 - \lambda, \\ 1 - \mathrm{K}, & \text{with probability } \lambda, \end{cases} \tag{31}$$

where $\beta_0 = \lambda_0 - \lambda$ and ι is a conformable vector of ones. Under the null hypothesis, of Equation (24), the regressors should have no explanatory power, that is $\mathrm{H}_0 : \boldsymbol{\beta} = \mathbf{0}$. Because the regressors are not correlated with the dependent variables under the null hypothesis, invoking a suitable central limit theorem (CLT) yields[15]

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{H} - \lambda\iota) \overset{\mathrm{asy}}{\sim} \mathrm{N}\left(\mathbf{0}, (\mathbf{X}'\mathbf{X})^{-1}\lambda(1 - \lambda)\right),$$

from which Engle and Manganelli (2004) deduce the test statistic

$$\mathrm{DQ} = \frac{\hat{\boldsymbol{\beta}}'_{\mathrm{LS}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}'_{\mathrm{LS}}}{\lambda(1 - \lambda)} \overset{\mathrm{asy}}{\sim} \chi^2_{p+2}. \tag{32}$$

In the empirical application below, we use two specifications of the dynamic quantile (DQ) test: For the first, denoted by $\mathrm{DQ}_{\mathrm{Hit}}$, regressor matrix $\mathbf{X}$ contains a constant and four lagged hits, $H_{t-1}, \ldots, H_{t-4}$; the second, $\mathrm{DQ}_{\mathrm{VaR}}$, uses, in addition, the contemporaneous VaR estimate.

3   EMPIRICAL ANALYSIS

We examine the VaR forecasting performance for a portfolio that is long in the NASDAQ Composite Index.[16] The index itself is a market value-weighted portfolio of more than 5000 stocks listed on the NASDAQ stock market. The data comprise daily closing levels, p_t, of the index from its inception on February 8, 1971, to June 22, 2001, yielding a total of $T = 7681$ observations of percentage log returns, $r_t := 100(\ln p_t - \ln p_{t-1})$. Table 1 presents some relevant summary statistics. The

---

[15] Note that for the asymptotics under the null hypothesis, it is irrelevant whether $H_t - \lambda$ is regressed on lags of H_t or lags of $H_t - \lambda$, as proposed by Engle and Manganelli (2004).

[16] The data and further information may be obtained from http://www.marketdata.nasdaq.com, maintained by the Economic Research Department of the National Association of Securities Dealers, Inc.

Table 1 Summary statistics for NASDAQ returns.

Sample Size	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max
7681	0.0392	1.1330	−0.4656	17.302	−12.048	13.255

sample skewness $\hat{\mu}_3/\hat{\mu}_2^{3/2} = -0.466$ indicates considerable asymmetry which, taken together with the sample kurtosis $\hat{\mu}_4/\hat{\mu}_2^2 = 17.3$, indicates a substantial violation of normality.[17]

3.1   Window Size 1000

For all models considered, we allow the corresponding parameters to change over time: Using moving windows initially of size $w = 1000$ (corresponding to roughly four years of trading data), we update the model parameters for each moving window with increments of one trading day. This leaves us with 6681 one-step-ahead VaR forecasts to study the predictive performance of the models. To save space, we refer to the AR(1)-GARCH(1,1) specification simply as "GARCH."[18]

In addition to the GARCH-EVT approach of Diebold, Schuermann, and Stroughair (1998) and McNeil and Frey (2000), which relies on the normal assumption for the GARCH filter, we consider an alternative conditional EVT implementation that specifies the skewed t instead of the normal distribution in order to better account for conditional asymmetry and heavy-tailedness. In Table 3 and the subsequent discussion, these are referred to as N-EVT and ST-EVT, respectively. We also entertain the FHS. The simulation itself is nonparametric, using standardized historical residuals from applying the normal and the skewed-t filter, marked as N-FHS and ST-FHS respectively, to the data.

The results under the alternative modeling assumptions are reported in Tables 2, 3, and 5. A plot of a selection of the VaR forecasts can be found in Figure 1. With a few exceptions, all models tend to underestimate the frequency of extreme returns.

---

[17] The fact that heavy-tailed distributions may not possess low-order moments implies that usual significance tests for skewness and kurtosis are most likely unreliable and are not worth reporting. See Loretan and Phillips (1994), Adler, Feldman, and Gallagher (1998), Paolella (2001), and the references therein for further discussion.

[18] The significant autocorrelation of returns is dying out toward the end of the sample. An anonymous referee has pointed out that the autocorrelation could be an artifact of stale quotes and suggested prefiltering the data. The results for the unconditional models did not improve when doing so. All other models entertained allow for autocorrelation (and hence appropriate filtering) except for the indirect GARCH(1,1) CAViaR, for which we provide an extension. Consequently we do not bias our results against any of the presented approaches. In addition, certain trading patterns may endogenously alter the correlation properties in index data in more volatile times. Venetis and Peel (2005), for instance, examine empirically whether there is an inverse volatility-correlation relationship in index returns and conclude in favor. To us, this would suggest that autocorrelation is endogenous and should be modeled jointly with volatility.

Table 2 VaR prediction performance: unconditional models.[a]

Model	100λ	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	$\overline{VaR}$
HS	1	1.30	**0.02**	0.00	0.00	0.00	0.00	2.65
	2.5	3.26	0.00	0.00	0.00	0.00	0.00	1.96
	5	6.00	0.00	0.00	0.00	0.00	0.00	1.43
Normal	1	2.80	0.00	0.00	0.00	0.00	0.00	2.06
	2.5	4.27	0.00	0.00	0.00	0.00	0.00	1.73
	5	6.18	0.00	0.00	0.00	0.00	0.00	1.44
t	1	2.10	0.00	0.00	0.00	0.00	0.00	2.30
	2.5	4.52	0.00	0.00	0.00	0.00	0.00	1.66
	5	7.74	0.00	0.00	0.00	0.00	0.00	1.25
Skewed t	1	1.30	**0.02**	0.00	0.00	0.00	0.00	2.64
	2.5	3.46	0.00	0.00	0.00	0.00	0.00	1.90
	5	6.17	0.00	0.00	0.00	0.00	0.00	1.41
EVT	1	1.29	**0.02**	0.00	0.00	0.00	0.00	2.65
	2.5	3.40	0.00	0.00	0.00	0.00	0.00	1.92
	5	**6.03**	0.00	0.00	0.00	0.00	0.00	1.43

[a]Results pertain to the 1000-length window size. HS is the naive historical simulation; λ is the target probability. Entries in the last five columns are the significance levels (p-values) of the respective tests. Bold type entries are not significant at the 1% level. For DQ_{Hit}, $H_t - \lambda$ is regressed onto a constant and four lagged violations, for DQ_{VaR}, in addition the contemporaneous VaR estimate. The unconditional EVT model does not "prefilter" with an estimated ARMA-GARCH structure and is instead directly applied to the (negative) return data. $\overline{VaR}$ denotes the average value of the VaR estimates.

Although the performance varies substantially across the modeling approaches as well as the distributional assumptions, some clear patterns emerge. We first discuss the performance of the unconditional models, then the GARCH-based, and finally the CAViaR models.

As the unconditional models do not account for volatility clustering, none of them is able to produce iid VaR violations, causing us to strongly reject independence of the H_t sequences for all unconditional models (see Table 2).[19] At the 1% λ-level, the naive historical simulation performs well with respect to violation frequencies, along with the skewed t and the unconditional EVT. The superior performance of the skewed-t distribution relative to the normal and t is due to the fact that it allows for skewness, which is clearly present in the unconditional return data (see Table 1).

Current regulatory performance assessment focuses on the unconditional coverage property, leaving other implications of efficiency unexamined. The three-zone framework suggested by the Basle Committee (1996b) deems a VaR model acceptable (green zone) if the number of violations of 1% VaR remains

---

[19] This finding contradicts Danielsson and de Vries (2000) and is instead more in line with the observations of Danielsson and Morimoto (2000), who—for their datasets—still observe considerable (though decreasing) dependence in extreme returns above (and including) the 1% λ-level.

Table 3 VaR prediction performance: AR(1)-GARCH(1,1).[a]

Model	100λ	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	$\overline{VaR}$
Normal	1.0	2.23	0.00	**0.03**	0.00	0.00	0.00	2.05
	2.5	3.92	0.00	**0.04**	0.00	0.00	0.00	1.72
	5.0	6.21	0.00	**0.21**	0.00	0.00	0.00	1.43
Student's t	1.0	1.81	0.00	0.01	0.00	0.00	0.00	2.19
	2.5	4.04	0.00	**0.02**	0.00	0.00	0.00	1.70
	5.0	6.89	0.00	**0.06**	0.00	0.00	0.00	1.34
Skewed-t	1.0	1.20	**0.12**	**0.35**	**0.19**	**0.16**	**0.04**	2.57
	2.5	2.72	**0.25**	0.00	0.00	0.00	0.00	2.01
	5.0	5.12	**0.65**	**0.03**	**0.08**	**0.07**	0.00	1.59
MixN(2,2)	1.0	0.91	**0.47**	**0.59**	**0.67**	**0.16**	**0.18**	2.53
	2.5	2.86	**0.07**	**0.04**	0.00	0.00	**0.01**	1.92
	5.0	5.78	0.00	**0.02**	0.00	0.00	0.00	1.45
MixN(3,2)	1.0	1.29	**0.02**	**0.13**	**0.02**	0.00	0.00	2.39
	2.5	2.86	**0.07**	**0.01**	0.00	0.00	0.00	1.88
	5.0	5.66	**0.02**	**0.06**	**0.01**	0.00	0.00	1.48
MixN(3,3)	1.0	1.18	**0.14**	**0.33**	**0.22**	0.00	0.00	2.46
	2.5	2.93	**0.03**	**0.04**	**0.01**	0.00	0.00	1.91
	5.0	5.55	**0.04**	**0.22**	**0.06**	0.00	0.00	1.50
MixGED(2,2)	1.0	1.06	**0.61**	**0.79**	**0.85**	**0.04**	**0.03**	2.53
	2.5	2.72	**0.25**	**0.10**	**0.12**	**0.02**	**0.03**	1.94
	5.0	5.37	**0.17**	**0.05**	**0.05**	0.00	0.00	1.50
MixGED(3,2)	1.0	1.14	**0.27**	**0.89**	**0.54**	**0.06**	**0.06**	2.52
	2.5	2.57	**0.70**	**0.05**	**0.13**	**0.05**	**0.07**	1.94
	5.0	5.10	**0.70**	**0.04**	**0.12**	0.00	0.00	1.51
MixGED(3,3)	1.0	1.23	**0.07**	**0.38**	**0.13**	**0.04**	0.01	2.58
	2.5	2.51	**0.94**	**0.01**	**0.05**	0.00	0.00	1.96
	5.0	5.16	**0.54**	**0.09**	**0.19**	0.00	0.00	1.52
N-EVT	1.0	0.97	**0.82**	**0.16**	**0.37**	**0.12**	**0.08**	2.61
	2.5	2.50	**1.00**	0.01	**0.02**	0.00	0.00	2.00
	5.0	5.33	**0.22**	**0.07**	**0.09**	**0.05**	**0.08**	1.54
ST-EVT	1.0	0.97	**0.82**	**0.17**	**0.37**	**0.09**	**0.02**	2.70
	2.5	2.47	**0.87**	0.00	0.00	0.00	0.00	2.07
	5.0	5.06	**0.82**	**0.02**	**0.06**	**0.08**	0.00	1.61
N-FHS	1.0	1.06	**0.60**	**0.05**	**0.12**	0.00	0.00	2.51
	2.5	2.76	**0.18**	0.00	0.00	0.00	0.00	1.94
	5.0	5.28	**0.30**	**0.05**	**0.09**	**0.04**	**0.07**	1.54
ST-FHS	1.0	0.94	**0.64**	**0.14**	**0.31**	**0.09**	**0.02**	2.71
	2.5	2.61	**0.57**	0.00	0.00	0.00	0.00	2.06
	5.0	4.96	**0.89**	**0.01**	**0.04**	**0.05**	0.00	1.61

[a]Results pertain to the 1000-length window size. MixN(k, g) and MixGED(k, g) refer to the MixN(k, g)-GARCH(1,1) and MixGED(k, g)-GARCH(1,1) models, respectively, in Section 1.3, with an AR(1) term for the mean. N-EVT refers to the use of AR(1)-GARCH(1,1) with Gaussian innovations as the filter used with the conditional EVT model; ST-EVT is similar, but uses the density of Equation (5) instead. N-FHS refers to the use of AR(1)-GARCH(1,1) with Gaussian innovations as the filter used with the conditional FHS model; ST-FHS is similar, but uses the density of Equation (5) instead. $\overline{VaR}$ denotes the average value of the VaR estimates. See also the note in Table 2.
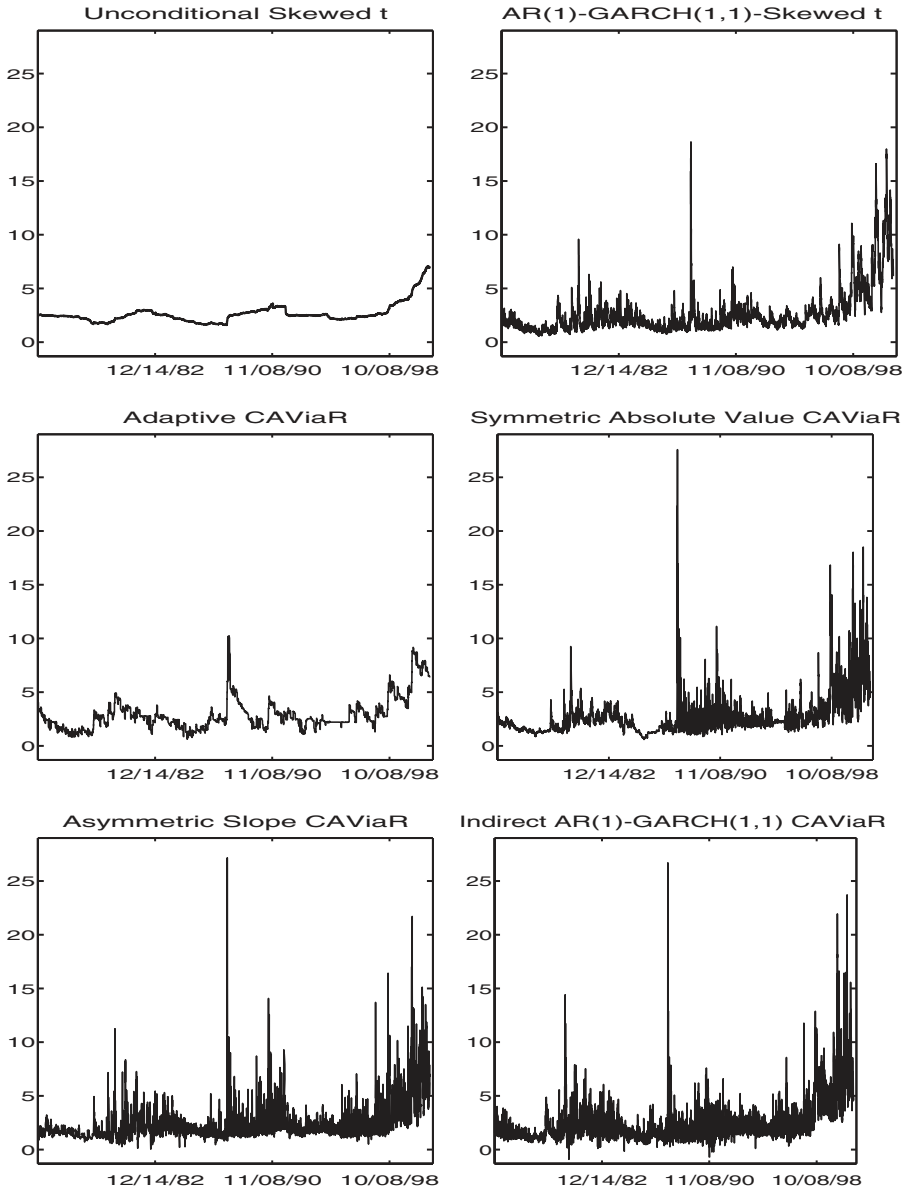
Figure 1 One-step ahead VaR predictions for some of the methods for $\lambda = 0.01$.

below the binomial(0.01) 95% quantile. A model is disputable (yellow zone) up to the 99.99% quantile and is deemed seriously flawed (red zone) whenever more violations occur. Translated to our sample size, a model passes regulatory performance assessment if, at most, 79 violations (1.18%) occur and is disputable when between 80 and 98 (1.47%) violations occur. The results reported in Table 2

suggest that, as far as the unconditional coverage is concerned, the normal and Student's *t* unconditional models are seriously flawed (red), while naive historical simulation, unconditional EVT, and the unconditional skewed-*t* models are disputable (yellow), though the latter are still inadequate in terms of clustering of violations. As none of the unconditional models is acceptable in any of the testing categories at both the 2.5% and 5% target probability, a conditional approach should be preferred.

Introducing GARCH volatility dynamics almost uniformly improves VaR prediction performance: The great majority of models do better than the unconditional approach across all λ-levels and all distributions considered—as was to be expected from the apparent volatility clustering in the return series. Regarding the percentage of violations (see Table 3), among the distributional assumptions for the fully parametric models, the skewed *t* is by far the best for all three λ-levels. Both the *t* and normal performed quite poorly. While this is not surprising for the normal, some empirical studies, such as Danielsson and Morimoto (2000) and McNeil and Frey (2000), show that normal GARCH might have some merit for larger values of λ. Our findings indicate that, at least for this dataset, the 5% quantile is still not large enough for the normal assumption to be adequate.

To clearly illustrate this point and expedite the comparison of VaR forecasting methods, we advocate use of a simple, but seemingly novel graphical depiction of the quality of VaR forecasts over all the relevant probability levels. This is shown in Figure 2, which depicts the coverage results for all VaR levels up to $\lambda = 0.1$. It plots the forecast cdf against the deviation from a uniform cdf. The VaR levels can be read off the horizontal axis, while the vertical axis depicts, for each VaR level, the excess of percentage violations over the VaR level. Thus the relative deviation from the correct coverage can be compared across VaR levels and competing models. One immediately spots that the normal assumption yields quite accurate results for λ around 0.08 (corresponding to "8" on the horizontal axis in the plot), in stark comparison to the Student's *t* assumption, which tends to perform worse as λ increases and yields the worst result among all the conditional models.

We next turn to the mixture models of Section 1.3. Their deviation plots are shown (using a less favorable scale than that used for the aforementioned models) in the bottom panels of Figure 2. It is apparent that the MixN(n, g) models exhibit good unconditional coverage properties at the low quantiles. The MixN(3,3) performs best overall, indicating (for this dataset and choice of window length) that three heteroskedastic mixture components are most suitable. Turning to the MixGED models, we see that, in terms of forecasting ability, they clearly stand out. All three MixGED models stay quite close to their prescribed violation rates for all VaR levels depicted, with preference going almost equally to the MixGED(3,2) and MixGED(3,3). The third component of the mixture corresponds to the lowest (i.e., most fat-tailed) GED density (see the discussion in the next paragraph), so that, when compared to the MixN findings, it appears that use of a fat-tailed component obviates use of the third heteroskedasticity structure. This
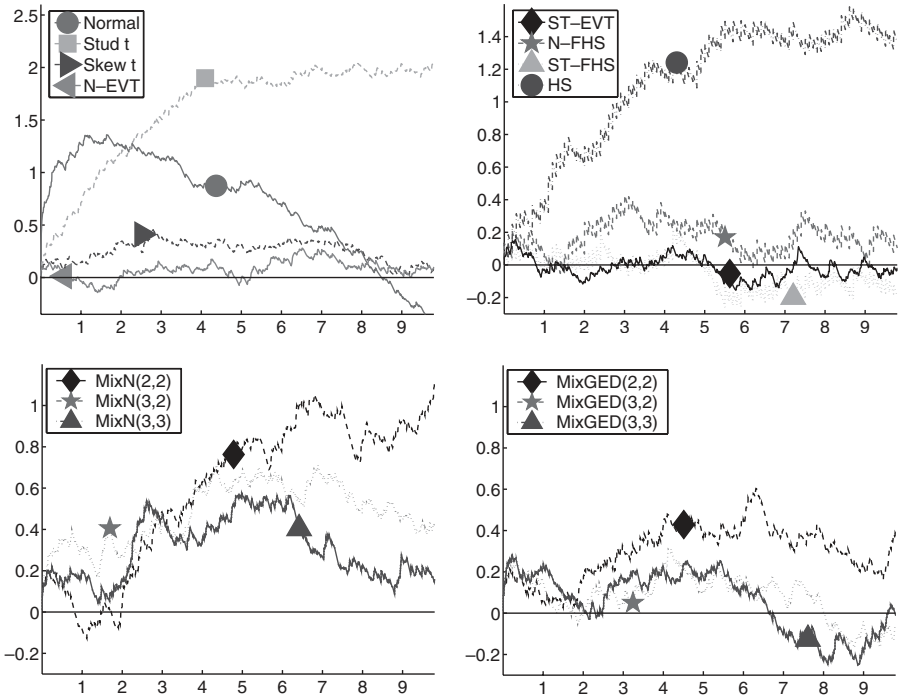
Figure 2 Deviation probability plot for the GARCH-based conditional models. Plotted values are "Deviation" := $100\left(F_U - \hat{F}\right)$ (vertical axis) versus $100\hat{F}$ (horizontal axis), where F_U is the cdf of a uniform random variable; $\hat{F}$ refers to the empirical cdf formed from evaluating the 6681 one-step, out-of-sample distribution forecasts at the true, observed return. Note the different scale in the first two plots. The upper right plot also shows the forecast density deviations for plain historical simulation (using data prefiltered to correct for first-order autocorrelation by ordinary least squares).

result is intuitive, recalling that both normal GARCH or an iid fat-tailed model exhibit excess kurtosis.

The improvement of the MixGED over the MixN obviously lies in the introduction of the GED exponent parameters, p_i. To get a better idea of their impact, Figure 3 plots their estimated values through the moving window of 6681 samples (of size $w = 1000$). They move less erratically than might be expected, and $\hat{p}_3$ barely deviates from its mean of 1.65. It thus appears that enhancing the MixN model with the flexibility to accommodate both fat- and thin-tailed innovations significantly enhances forecasting performance. The benefit of the thin-tailed aspect allowed for in the MixGED is limited, however. Figure 4 shows the standard normal density ($p = 2$) and the GED density corresponding to $p = 2.10$, $p = 1.72$, and $p = 1.66$, which are the estimated values of the GED shape parameter of the last 1000-length sample, as shown in Figure 3. One sees that the difference between GED densities with $p = 2$ and $p = 2.10$ is too slight to intuitively judge as
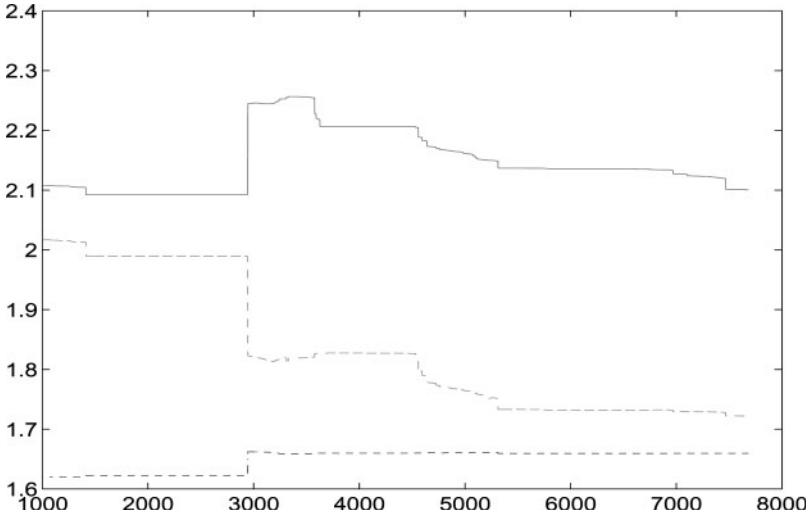
Figure 3 The estimated values of the GED exponent parameter p_i, $i = 1, 2, 3$, in the MixGED(3, 2) model, through the moving window of 6681 samples of size 1000. The solid line is $\hat{p}_1$, dashed is $\hat{p}_2$, and dash-dot is $\hat{p}_3$.
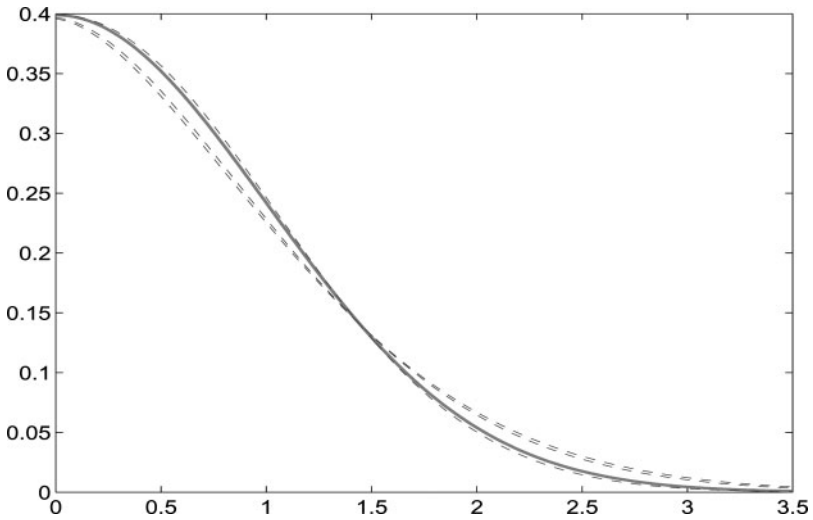
Figure 4 The normal density function (thick solid line) and the GED density function for $p = 2.10$, $p = 1.72$, and $p = 1.66$ (dashed), which correspond to the estimated values of the GED shape parameter p of the last 1000-length sample, as shown in Figure 3. The GED density is given in Equation (8), but is scaled by $\sqrt{2}$ so that it coincides with the standard normal density for $p = 2$.

significant.[20] Perhaps the best way of confirming this is by computing all three entertained MixGED(n, g) models, but restricting each p_i to be bounded above by 2. This was done and was found to yield very similar results to those in Figure 2 (results available upon request).

Like the mixture models, the EVT-based approaches outperform the fully parametric conditional location-scale models, discussed in Section 1.2, with pre-ference given to the ST-EVT. In particular, from Table 3, the N-EVT has 0.97%, 2.50%, and 5.33% violations for $\lambda = 0.01$, 0.025, and 0.05, respectively, and ST-EVT has 0.97%, 2.47%, and 5.06%. Similar results emerge for the FHS approaches. This is also made clear in Figure 2: The two EVT formulations and the FHS approaches are always among the best performers among all conditional models and for all values of $0 < \lambda < 0.1$.[21] Moreover, the ST-EVT delivers virtually exact results for all $0 < \lambda < 0.1$, while N-EVT is competitive for $0 < \lambda$ 0.025 and then weakens somewhat as λ increases toward 0.1.

Table 4 provides summary information for each model by averaging over all values of λ in (0,0.1). To construct a measure of fit, we computed the mean absolute error (MAE) and mean-squared error (MSE) of the actual violation frequencies from the corresponding theoretical VaR level (based on the first 660 of the ordered deviations shown in Figure 2). Both summary measures indicate the same result: the ST-EVT has (with considerable margin) the lowest deviation over the entire tail.

We now turn to the information in the sequence of violations, as reflected in the p values of the LR and DQ test statistics in Table 3. Entries in bold type are greater than 0.01, signifying that the hypothesis of independence cannot be rejected at the 1% significance level. Eleven models exhibit all p values greater than 0.01, thus hinting at efficient VaR forecasts: the fully parametric GARCH with skewed-t innovations and the MixN (2,2) for $\lambda = 0.01$; the MixGED(2,2) and MixGED(3,2) for $\lambda = 0.01$ and $\lambda = 0.025$; the N-EVT for $\lambda = 0.01$ and 0.05; the ST-EVT for $\lambda = 0.01$; the N-FHS for $\lambda = 0.05$; and the ST-FHS for $\lambda = 0.01$. From this list and the other tabulated p values of the LR and DQ statistics, it is clear overall that, with respect to independence, when the AR(1)-GARCH(1,1) filter with a skewed, fat-tailed distributional assumption is combined with the FHS or the EVT model, VaR violations result which contain virtually no information about the probability of a future violation.[22] The same applies to AR(1)-GARCH(1,1) models combined

---

[20] Ideally 95% one-at-a-time confidence intervals would accompany the point estimates shown in Figure 3. However, the error bounds we obtained based on the numerically computed Hessian matrix (as implemented in Matlab's optimization routines) were erratic and sometimes implausible. A bootstrap exercise could be used to get reliable standard errors, but this would be required for each moving window, and so would be impossible in terms of computing time.

[21] The FHS forecast densities seem to oscillate. This is due to the cdf of the FHS being a step function, and is therefore an artifact of the method.

[22] Note that the violation frequencies for the AR(1)-GARCH(1,1) filter with either the normal or t innova-tion assumption are considerably worse than for the S&P and DAX series reported in McNeil and Frey (2000). However, none of their tests includes the high-volatility regime following the Asian and Russian crises, as well as the beginning of the recent slump in the market.

Table 4 Overall deviation measures.[a]

| | Window size | | | | | |
| | 1000 | | 500 | | 250 | |
GARCH model	MAE	MSE	MAE	MSE	MAE	MSE
Normal	0.970	1.147	0.770	0.773	0.785	0.809
Student's t	1.474	2.504	1.309	1.966	1.376	2.099
Skewed t	0.146	0.031	0.156	0.030	0.263	0.089
MixN(2,2)	0.565	0.429	0.655	0.566	0.503	0.266
MixN(3,2)	0.470	0.244	0.243	0.080	0.196	0.045
MixN(3,3)	0.304	0.115	0.764	0.702	0.570	0.356
MixGED(2,2)	0.292	0.102	0.381	0.170	0.418	0.180
MixGED(3,2)	0.120	0.018	0.125	0.021	0.450	0.214
MixGED(3,3)	0.140	0.024	0.860	0.813	0.555	0.341
N-EVT	0.217	0.065	0.091	0.011	0.404	0.179
ST-EVT	0.050	0.004	0.188	0.050	0.197	0.049
N-FHS	0.174	0.039	0.101	0.015	0.461	0.239
ST-FHS	0.093	0.013	0.164	0.033	0.206	0.059
FHS	0.999	1.173	0.627	0.431	0.449	0.228

[a]The mean of the absolute (MAE) and squared (MSE) error of the empirical from the theoretical tail probability ("Deviation" in Figure 2). This is computed over the first 9.9% of the sorted out-of-sample cdf values for the GARCH-based models.

with flexible mixture distributions. In particular, the GED mixture class of models is the only one that achieves passing LR tests at all three VaR levels. Of the models that perform reasonable overall (the skewed-t, MixGED, FHS, and EVT models), the N-FHS and MixGED models have the lowest average VaR across all VaR levels. In other words, N-FHS and MixGED on average bring about the lowest regulatory capital requirement—followed by skewed-t GARCH and N-EVT.

Summarizing the results for the fully parametric, FHS, and EVT models: Major improvements in VaR predictions are achieved in all aspects when accounting for the volatility dynamics. VaR violations are reasonably independent when using either the fully parametric GARCH model with skewed-t innovations or a mixture model, the FHS, or an EVT model based on a GARCH filter with either normal or skewed-t innovations—the latter and the GED mixtures being preferred overall.

Next, we turn to the results of the CAViaR models, which deliver mixed results (see Table 5). Only the very simple adaptive CAViaR specification performs adequately at all λ-levels with regard to unconditional coverage. Because the adaptive CAViaR model increases VaR once a violation occurs and decreases it slightly otherwise, it is not surprising that it cannot produce cluster-free violations. This also agrees with the results reported in Engle and Manganelli (2004).

The symmetric absolute value specification, which adapts VaR to the size of returns, is only acceptable with respect to unconditional coverage at the higher

Table 5 VaR prediction performance: CAViaR.[a]

Model	100λ	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	$\overline{VaR}$
Adaptive	1	1.14	**0.27**	0.00	0.00	0.00	0.00	2.88
	2.5	2.80	**0.12**	0.00	0.00	0.00	0.00	2.13
	5	5.10	**0.70**	0.00	0.00	0.00	0.00	1.66
Symmetric abs. value	1	1.33	0.01	0.01	0.00	0.00	0.00	2.69
	2.5	2.83	**0.09**	0.00	0.00	0.00	0.00	2.03
	5	5.45	**0.10**	0.00	0.00	0.00	0.00	1.61
Aymmetric slope	1	1.60	0.00	**0.83**	0.00	0.00	0.00	2.39
	2.5	3.35	0.00	**0.01**	0.00	0.00	0.00	1.85
	5	6.02	0.00	**0.11**	0.00	0.00	0.00	1.48
Indirect GARCH(1,1)	1	1.32	**0.01**	0.00	0.00	0.00	0.00	2.67
	2.5	3.08	0.00	0.00	0.00	0.00	0.00	2.03
	5	5.54	**0.05**	0.00	0.00	0.00	0.00	1.59
Indirect AR(1)-GARCH(1,1)	1	1.32	**0.01**	0.15	0.02	0.00	0.00	2.49
	2.5	3.04	0.00	**0.74**	0.02	0.03	0.01	1.91
	5	5.55	**0.04**	0.89	0.12	0.01	0.02	1.49

[a]Results pertain to the 1000-length window size. See the note in Table 2 regarding the statistical tests and measures, and Section 1.5 for a description of the models.

λ-levels. Otherwise it blatantly fails the independence tests, rendering its performance as inadequate as that of the adaptive CAViaR. While the asymmetric slope specification is a generalization of the symmetric absolute value CAViaR, it exhibits quite different characteristics: It passes the independence test at all λ-levels, but it fails all the other tests at all λ levels.

None of the traditional CAViaR models passes any DQ test at any λ-level. The indirect GARCH(1, 1) specification performs well at the 1% and 5% λ-levels, yet only with respect to correct unconditional coverage. This was expected, given the first-order autocorrelation in the data. Engle and Manganelli (2004) report a weaker CAViaR performance for index data than for individual stock returns. The fact that our NASDAQ sample comprises two additional highly volatile years, which presumably deteriorate overall performance, may help to reconcile the poor out-of-sample performance of the established CAViaR models with the more positive findings of Engle and Manganelli (2004).

The need to incorporate an explicit autoregressive link to returns is seen when looking at the indirect CAViaR AR(1)-GARCH(1, 1) model proposed here.[23] It passes 12 out of the 15 tests, whereas the second best CAViaR specification passes only 3 out of the 15. While resulting in too many VaR violations for the 2.5% level, the model exhibits correct unconditional coverage at the 1% and 5% λ-levels. All other CAViaR models fail in all DQ tests, yet the proposed specification

---

[23] As regards autocorrelation and the degree to which this may only be an artifact of the index data examined here, we refer to note 18.

passes for the 2.5% and 5% VaR levels. Hence, in the sense that it is less prone to violation clustering, the new model improves considerably on the previous CAViaR specifications.

3.2 Smaller Window Sizes

So far, the analysis has been based on the use of a moving window of $w = 1000$ observations, but it can be decisive to know the extent to which the results are compatible when using different, particularly shorter, lengths. To address this, we repeat all the calculations using $w = 500$ and $w = 250$, except for the CAViaR models, given their poorer performance. To save space, we consider only their coverage performance; the deviation plots in Figure 5 show these for $w = 500$, while Table 4 includes the summary deviation measures for both cases.

Some very clear conclusions easily emerge from comparison of Figures 2 ($w = 1000$) and 5 ($w = 500$). The normal and Student's t GARCH models again perform very poorly, exhibiting much the same "shape" in the deviation plot for both values of w, while the skewed-t GARCH again does remarkably well. Both EVT models also perform well, though the relative performance of N-EVT and ST-EVT has reversed when compared to the $w = 1000$ case. This cautiously signals
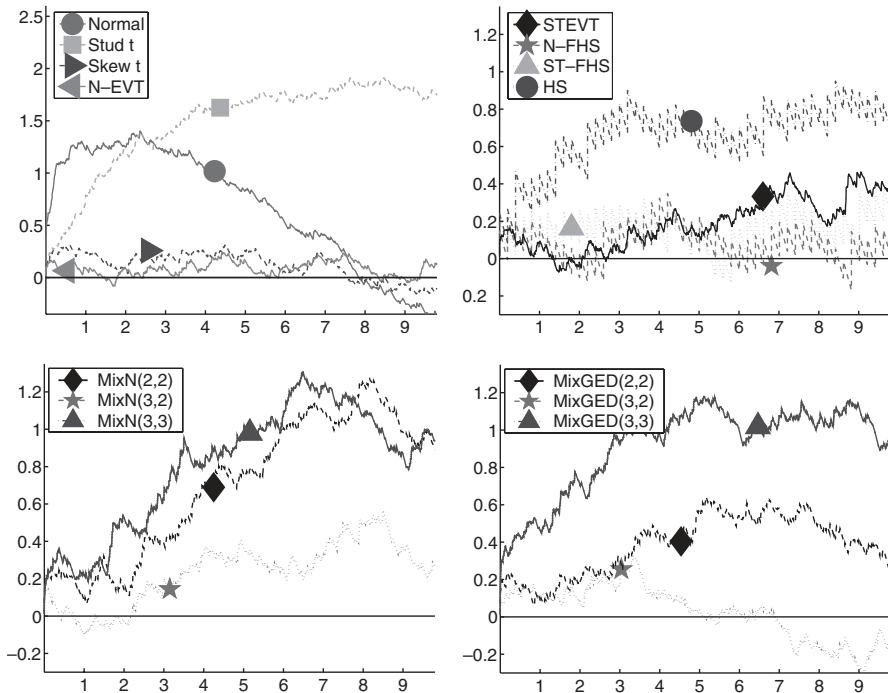

Figure 5 Same as Figure 2, but based on a window size of 500.

that a trade-off exists between sample size and model robustness, with smaller samples benefiting from a simpler GARCH filter. Regarding the FHS methods, both again perform admirably well, indicating their applicability for a range of sample sizes. As a point of reference, in the upper left panel we also provide the deviations of the historical simulation for this sample size.[24] Notably, a number of the parametric methods presented here considerably outperform the plain historical simulation even as the sample size shrinks.

The performance of the mixture models is precisely in line with our expectations. While still far better than the normal and t GARCH models, overall the MixN models perform worse with $w = 500$, due presumably to their rather large parameterization, though the MixN(3, 2) performs reasonably well. We conjecture that the MixN(2, 2) case is inadequate for capturing the dynamics of the series, but the MixN(3, 3) case is overly parameterized with respect to this sample size. This is also supported by the MixGED results, for which the MixGED(3, 2) is among the top performers of all the entertained models, but the (2, 2) and, particularly, the (3, 3) cases do poorly.

The window size of $w = 250$ should be the most challenging for all models, though from Table 4 we see that the MAE and MSE actually decrease when moving from $w = 1000$ to $w = 500$ to $w = 250$ for some models, such as MixN(3, 2), and general conclusions about model performance as the window length decreases cannot be made. Comparisons across models for the $w = 250$ case are possible, however, and are quite in agreement with the results for larger w. In particular, we see that the skewed-t GARCH model is vastly superior to its normal or symmetric t counterparts, though it is, in turn, outperformed by the MixN(3, 2) model. Among the MixN models, the situation is the same as for the $w = 500$ case, that is, three components are necessary to capture the dynamics, but the third component does not need a GARCH specification for this small window size. Differences to the $w = 500$ case appear with the MixGED models, which are not competitive for $w = 250$ because of their relatively large parameterization. For the EVT models, the ST-EVT outperforms the N-EVT, as was the case for $w = 1000$, again lending evidence of the importance of appropriate distributional assumptions in this model class. The best models for $w = 250$ are the ST-EVT, MixN(3, 2), and ST-FHS, which all perform virtually identically with regard to the MAE and MSE measures over $\lambda \in (0, 0.1)$.

4   CONCLUSION

The predictive performance of several recently advanced and some new VaR models has been examined. The majority of these suffer from excessive VaR violations, implying an underestimation of market risk. Employing more informative tests than is common in the literature, we find that regulatory forecasting

---

[24] For the results shown we have prefiltered the data using ordinary least squares to remove first-order autocorrelation. The plot for unconditional historical simulation on the raw data, however, looks quantitatively very similar.

assessment can be flawed. Most notably, all of the unconditional models produce clustered VaR violations, yet some may still pass as acceptable when considering only the (unconditional) violation frequencies.

Conditional VaR models lead to much more volatile VaR predictions than unconditional models and may arguably cause problems in allocating capital for trading purposes [see, e.g., Danielsson and Morimoto (2000)]. However, our results show that only conditionally heteroskedastic models yield acceptable forecasts. For the fully parametric models, a major improvement in terms of violation frequencies is achieved when accounting for scale dynamics. In addition, taking heteroskedasticity into account yields reasonably unclustered VaR violations. Considerable improvement in normality is achieved by using innovation distributions that allow for skewness and fat tails. The conditional skewed-t, MixGED(n, g), two EVT (N-EVT and ST-EVT), and two FHS approaches (N-FHS and ST-FHS) perform best, though this conclusion depends to some extent on the chosen window size, with less-parameterized models having an advantage as w decreases from 1000 to 250.

It is worth noting the presence of the skewed t among the best models (skewed-t GARCH, ST-EVT, and ST-FHS), adding further evidence to the findings of work cited in Section 1.2. It appears that the FHS methods are the most robust to the choice of window length, though the other top performers did not suffer much when changing w. The finding regarding the importance of the skewed t in the EVT and FHS models is interesting, if not disturbing, because it implies that distributionally nonparametric models do indeed depend on the distribution assumed in the filtering stage.

Finally, none of the CAViaR models perform well overall, though the proposed indirect AR(1)-GARCH(1, 1) model is most promising in the CAViaR class, in the sense that it passes most of the tests. In a dataset without autocorrelation, or in prefiltered data, the indirect GARCH(1, 1) CAViaR specification therefore is expected to also yield reasonable results.

Depending on the application, multistep VaR forecasting may be necessary. Due to the nonlinearity inherent in the GARCH dynamics, multistep-ahead forecasts require simulation. In particular, one would draw a large number of sequences of shocks from either the estimated innovations distribution or directly from the filtered residuals, and use these to simulate return paths. The resulting distribution can then be used to compute the desired VaR forecasts.

For the CAViaR models, the drawback in forecasting is that no return process is estimated along with the quantile process. For multistep-ahead forecasts, one could, in principle, separately estimate a model for the return process and a CAViaR model for the quantile. The model for the return process could then be used to simulate paths of return series and the CAViaR model would consequently deliver forecasts for the VaRs. A multistep-ahead forecast could, for example, rely on the mean VaR. Alternatively, one could understand the CAViaR models as representing h-day holding returns. The models would then need to be reestimated using multiday returns.

In this article we restrict ourselves to one-step-ahead forecasts, leaving the more general case for future work.

APPENDIX

We implemented the EVT approach unconditionally, using the raw return data, and conditionally (i.e., GARCH filtered) assuming normal and skewed-t innovations. In simulations of size 1000, for t-residuals, McNeil and Frey (2000) found that $k \approx 100$ minimizes the MSE of the resulting quantile estimates $\hat{z}'_{1-\lambda,k}$ and that the results are rather insensitive to the choice of k for a wide range of k. While this choice may not be adequate for other innovation distributions, there exist no automatic mechanisms capable of choosing k consistently [though see Gonzalo and Olmo (2004)]. We therefore followed their choice. Differentiating Equation (11) and assuming excesses are iid GPD, the log-likelihood function is

$$\mathcal{L}(\xi,\beta; y_1,\ldots,y_k) = \begin{cases} -k\log\beta - \left(\frac{1}{\xi}+1\right)\sum_{j=1}^{k}\log\left(1+\frac{\xi}{\beta}y_j\right), & \text{if } \xi \neq 0, \\ -k\log\beta - \sum_{j=1}^{k}\frac{y_j}{\beta}, & \text{if } \xi = 0, \end{cases}$$

with support $y_j \geq 0$, if $\xi \geq 0$, and with $0 \leq y_j \leq -\beta/\xi$, if $\xi < 0$. Smith (1987) showed that maximum-likelihood estimation works well for $\xi > -1/2$, a constraint that was never violated in our empirical analysis. Moreover, for iid data, we have

$$\sqrt{n}\left(\hat{\xi}-\xi, \frac{\hat{\beta}_n}{\beta}-1\right) \xrightarrow{d} N\left(0, M^{-1}\right), \quad n \to \infty, \tag{33}$$

where

$$M^{-1} = (1+\xi)\begin{bmatrix} 1+\xi & 1 \\ 1 & 2 \end{bmatrix}, \tag{34}$$

and the usual consistency and efficiency properties of maximum-likelihood estimation apply [Smith (1987)].

The support constraint for $\xi < 0$ is not easily implemented in standard constrained optimization. We chose to penalize the likelihood function proportionally to the amount of violation in order to "guide" the estimates back into regions covered by the GPD. We used the Nelder-Mead simplex algorithm, as implemented in the Matlab 6 routine "fminsearch," employing the default options.

Our implementation of the CAViaR approach was essentially the same as that in Engle and Manganelli (2004). Here, we discuss some relevant issues. Due to the multitude of local minima in the objective function of Equation (19), the baseline CAViaR model was estimated using a genetic algorithm [for details on the algorithm, see Engle and Manganelli (2004)]. The initial population consisted of 500 members chosen from the interval [0, 1]. A population size of 20 was maintained in each of a maximum of 500 generations. The programs used for calculation in Engle and Manganelli (2004) were unconstrained, which, for the NASDAQ data

we considered, resulted occasionally in negative estimates for parameter $\hat{\beta}$. In this case, each violation lowers the VaR measure. If toward the end of the in-sample period, the VaR measure is sufficiently low, the number of violations tends to accelerate, rapidly driving down the out-of-sample VaR forecasts to levels well below zero (toward $-\infty$ for long forecast horizons). This drawback may be overcome by decreasing the "genetic fitness" if this happens. We "punish" the fit by adding a value of three (about one-tenth of the fitness function for the 1% λ-level and correspondingly less for the higher levels) whenever VaR or the parameter estimate is negative. This proved to be successful in guiding the process back into more suitable regions. To estimate the remaining specifications, we used the Matlab 6 functions "fminunc" and "fminsearch," as described in Engle and Manganelli (2004).

The indirect GARCH models need to be constrained to prevent the terms in parentheses in Equations (20) and (23) from becoming negative. This was achieved by setting the objective function to 10^{10} whenever this occurred.

REFERENCES

Adler, R. J., R. E. Feldman, and C. Gallagher. (1998). "Analysing Stable Time Series." In R. J. Adler, R. E. Feldman, and M. S. Taqqu, (eds.), *A Practical Guide to Heavy Tails*. Boston: Birkhäuser.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. (1999). "Coherent Measures of Risk." *Mathematical Finance 9*:203–228.

Bao, Y., T.-H. Lee, and B. Saltoglu. (2003). "A Test for Density Forecast Comparison with Applications to Risk Management." Technical report, University of California, Riverside.

Bao, Y., T.-H. Lee, and B. Saltoglu. (2004). "Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check." Technical report, University of California, Riverside.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper. (1999). "VaR without Correlations for Portfolios of Derivative Securities." *Journal of Futures Markets 19*:583–602.

Barone-Adesi, G., K. Giannopoulos, and L. Vosper. (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." Available at http://www. bis.org.

Basle Committee on Banking Supervision. (1996a). "Overview of the Amendment to the Capital Accord to Incorporate Market Risks." Available at http://www. bis.org.

Basle Committee on Banking Supervision. (1996b). "Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." Available at http://www.bis.org.

Bassett, G., and R. Koenker. (1978). "Asymptotic Theory of the Least Absolute Error Regression." *Journal of the American Statistical Association 73*:618–622.

Bawa, V. S. (1978). "Safety-First, Stochastic Dominance, and Optimal Portfolio Choice." *Journal of Financial and Quantitative Analysis 13*:255–271.

Berkowitz, J., and J. O'Brien. (2002). "How Accurate Are Value-at-Risk Models at Commercial Banks?" *Journal of Finance LVII*:1093–1111.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics 31*:307–327.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge. (1988). "A Capital Asset Pricing Model with Time-Varying Covariances." *Journal of Political Economy 96*:116–131.

Bollerslev, T., and J. M. Wooldridge. (1992). "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances." *Econometric Reviews 11*:143–172.

Brooks, C., A. D. Clare, J. W. Dalle Molle, and G. Persand. (2005). "A Comparison of Extreme Value Theory Approaches for Determining Value at Risk." *Journal of Empirical Finance 12*:339–352.

Brooks, C., and G. Persand. (2003). "Volatility Forecasting for Risk Management." *Journal of Forecasting 22*:1–22.

Calvet, L., and A. Fisher. (2004). "Regime Switching and the Estimation of Multifractal Processes." *Journal of Financial Econometrics 2*:49–83.

Chernozhukoy, V., and L. Umantsev. (2001). "Conditional Value-at-Risk: Aspects of Modeling and Estimation." *Empirical Economics 26*:271–292.

Christoffersen, P. (1998). "Evaluating Interval Forecasts." *International Economic Review 39*:841–862.

Christoffersen, P. F. (2003). *Elements of Financial Risk Management*. Amsterdam: Academic Press.

Danielsson, J., and C. G. de Vries. (2000). "Value-at-Risk and Extreme Returns." *Annales d'Economie et de Statistique 60*:239–270.

Danielsson, J., and Y. Morimoto. (2000). "Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market." *Monetary and Economic Studies 18(2)*:25–48.

Diebold, F. X., T. Schuermann, and J. D. Stroughair. (1998). "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management." Working Paper 98–10, Wharton School, University of Pennsylvania.

Ding, Z., C. W. Granger, and R. F. Engle. (1993). "A Long Memory Property of Stock Market Returns and a New Model." *Journal of Empirical Finance 1*:83–106.

Dowd, K. (2002). *Measuring Market Risk*. Chichester: John Wiley & Sons.

Durbin, J., and S. J. Koopman. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Embrechts, P., C. Klüppelberg, and T. Mikosch. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Engle, R. F., and S. Manganelli. (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." *Journal of Business and Economic Statistics 22*:367–381.

Galbraith, J. W., and T. Kisinbay. (2005). "Content Horizons for Conditional Variance Forecasts." *International Journal of Forecasting 21*:249–260.

Gallati, R. R. (2003). *Risk Management and Capital Adequacy*. New York: McGraw-Hill.

Giot, P., and S. Laurent. (2004). "Modelling Daily Value-at-Risk Using Realized Volatility and ARCH Type Models." *Journal of Empirical Finance 11*:379–398.

Gonzalo, J., and J. Olmo. (2004). "Which Extreme Values are Really Extremes." *Journal of Financial Econometrics 2*:349–369.

Haas, M., S. Mittnik, and M. S. Paolella. (2004a). "Mixed Normal Conditional Heteroskedasticity." *Journal of Financial Econometrics 2*:211–250.

Haas, M., S. Mittnik, and M. S. Paolella. (2004b). "A New Approach to Markov Switching GARCH Models." *Journal of Financial Econometrics* 2:493–530.

Hansen, P. R., and A. Lunde. (2004). "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH (1,1)?" Working paper, Department of Economics, Brown University; forthcoming in *Journal of Applied Econometrics*.

Harvey, C. R., and A. Siddique. (1999). "Autoregressive Conditional Skewness." *Journal of Financial and Quantitative Analysis* 34:465–487.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk*. New York: McGraw-Hill.

Koenker, R., and G. Bassett. (1978). "Regression Quantiles." *Econometrica* 46:33–50.

Koenker, R., and S. Portnoy. (1997). "Quantile Regression." Working Paper 97-0100, University of Illinois at Urbana-Champaign.

Koenker, R., and Q. Zhao. (1996). "Conditional Quantile Estimation and Inference for ARCH Models." *Econometric Theory* 12:793–813.

Koopman, S. J., B. Jungbacker, and E. Hol. (2005). "Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realized and Implied Volatility Measurements." *Journal of Empirical Finance* 12:445–475.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. New York: John Wiley & Sons.

Ljung, G., and G. Box. (1979). "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 66:265–270.

Lopez, J. A. (1997). "Regulatory Evaluation of Value-at-Risk Models." Staff Report 33, November 1997, Federal Reserve Bank of New York.

Loretan, M., and P. Phillips. (1994). "Testing the Covariance Stationarity of Heavy-Tailed Time Series." *Journal of Empirical Finance* 1:211–248.

Lux, T. (2004). "The Markov-Switching Multi-Fractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility." Working paper, Christian Albrechts University, Kiel, Germany.

Lux, T., and T. Kaizoji. (2004). "Forecasting Volatility and Volume in the Tokyo Stock Market: The Advantage of Long Memory Models." Working paper, Christian Albrechts University, Kiel, Germany.

Martens, M. (2001). "Forecasting Daily Exchange Rate Volatility Using Intraday Returns." *Journal of International Money and Finance* 20:1–23.

McNeil, A. J., and R. Frey. (2000). "Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach." *Journal of Empirical Finance* 7:271–300.

Mittnik, S., and M. S. Paolella. (2000). "Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates." *Journal of Forecasting* 19:313–333.

Mittnik, S., and M. S. Paolella. (2003). "Prediction of Financial Downside-Risk with Heavy-Tailed Conditional Distributions." In S. T. Rachev (ed.), *Handbook of Heavy Tailed Distributions in Finance*. Amsterdam: North-Holland.

Nelson, D. B., and D. P. Foster. (1994). "Asymptotic Filtering Theory for Univariate ARCH Models." *Econometrica* 62:1–41.

Palm, F. C. (1996). "GARCH Models of Volatility." In G. S. Maddala and C. R. Rao (eds.), Amsterdam: North-Holland. *Handbook of Statistics: Statistical Methods in Finance*, vol. 14.

Paolella, M. S. (2001). "Testing the Stable Paretian Assumption." *Mathematical and Computer Modelling* 34:1095–1112.

Pedersen, C. S., and S. E. Satchell. (1998). "An Extended Family of Financial-Risk Measures." *Geneva Papers on Risk and Insurance Theory* 23(2):89–117.

Pickands, J. III. (1975). "Statistical Inference Using Extreme Order Statistics." *Annals of Statistics 3*:119–131.

Poon, S.-H., and C. Granger. (2003). "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature 41*:478–539.

Pritsker, M. (1997). "Evaluating Value at Risk Methodologies: Accuracy versus Computational Time." *Journal of Financial Services Research 12*:201–242.

Pritsker, M. (2001). "The Hidden Dangers of Historical Simulation." Finance and Economics Discussion Series 27. Board of Governors of the Federal Reserve System, Washington, D. C.

Rockinger, M., and E. Jondeau. (2002). "Entropy Densities with an Application to Autoregressive Conditional Skewness and Kurtosis." *Journal of Econometrics 106*:119–142.

Shephard, N., ed. (2005). *Stochastic Volatility*. Oxford: Oxford University Press.

Smith, R. L. (1987). "Estimating Tails of Probability Distributions." *Annals of Statistics 15*:1174–1207.

Taylor, J. W. (1999). "A Quantile Regression Approach to Estimating the Distribution of Multiperiod Returns." *Journal of Derivatives 7*:64–78.

Venetis, I. A., and D. Peel. (2005). "Non-Linearity in Stock Index Returns: The Volatility and Serial Correlation Relationship." *Economic Modelling 22*:1–19.

Vilasuso, J. (2002). "Forecasting Exchange Rate Volatility." *Economics Letters 76*:59–64.