Studies in Nonlinear Dynamics & Econometrics

Volume 10, *Issue* 2 | 2006 | *Article* 4

Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamics and Extreme Value Theory

Carlos Martins-Filho* Feng Yao†

*Oregon State University, carlos.martins@oregonstate.edu
†University of North Dakota, feng.yao@mail.business.und.edu

Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamics and Extreme Value Theory*

Carlos Martins-Filho and Feng Yao

Abstract

We propose an estimation procedure for value-at-risk (VaR) and expected shortfall (TailVaR) for conditional distributions of a time series of returns on a financial asset. Our approach combines a local polynomial estimator of conditional mean and volatility functions in a conditional heterocedastic autoregressive nonlinear (CHARN) model with Extreme Value Theory for estimating quantiles of the conditional distribution. We investigate the finite sample properties of our method and contrast them with alternatives, including the method recently proposed by McNeil and Frey (2000), in an extensive Monte Carlo study. The method we propose outperforms the estimators currently available in the literature. An evaluation based on backtesting was also performed.

*We would like to thank two anonymous referees for comments and suggestions. Department of Economics, Ballard Hall 303, Oregon State University, Corvallis, OR 97331-3612.

1 Introduction

The measurement of market risk to which financial institutions are exposed has become an important instrument for market regulators, portfolio managers and for internal risk control. As evidence of this growing importance, the Bank of International Settlements (Basle Committee, 1996) has imposed capital adequacy requirements on financial institutions that are based on measurements of market risk. Furthermore, there has been a proliferation of risk measurement tools and methodologies in financial markets (Risk, 1999). Two quantitative and synthetic measures of market risk have emerged in the financial literature, Value-at-Risk or VaR (RiskMetrics, 1995) and Expected Shortfall or TailVaR (Artzner et al., 1999). From a statistical perspective these risk measures have straightforward definitions. Let $\{Y_t\}$ be a stochastic process representing returns on a given portfolio, stock, bond or market index, where t indexes a discrete measure of time and F_t denotes either the marginal or the conditional distribution (normally conditioned on the lag history $\{Y_{t-k}\}_{M \geq k \geq 1}$, for some $M = 1, 2, ...$) of Y_t. For $0 < \alpha < 1$, the α-VaR of Y_t is simply the α-quantile associated with F_t.[1] Expected shortfall is defined as $E_{F_t^y}(Y_t)$ where the expectation is taken with respect to F_t^y, the truncated distribution associated with $Y_t > y$ where y is a specified threshold level. When the threshold y is taken to be α-VaR, then we refer to α-TailVaR.

Accurate estimation of VaR and TailVaR depends crucially on the ability to estimate the tails of the probability density function f_t associated with F_t. Conceptually, this can be accomplished in two distinct ways: a) direct estimation of f_t, or b) indirectly through a suitably defined (parametric) model for the tails of f_t. Unless estimation is based on a correct specification of f_t (up to a finite set of parameters), direct estimation will most likely provide a poor fit for its tails, since most observed data will likely take values away from the tail region of f_t (Diebold et al., 1998). As a result, a series of indirect estimation methods based on Extreme Value Theory (EVT) has recently emerged, including Embrechts et al. (1999), Longin (2000) and McNeil and Frey (2000). These indirect methods are based on approximating *only* the tails of f_t by an appropriately defined parametric density function.

In the case where f_t is a conditional density associated with a stochastic process of returns on a financial asset, a particularly promising approach is the two stage estimation procedure for conditional VaR and TailVaR suggested by McNeil and Frey. They envision a stochastic process whose evolution can be described as,

$$Y_t = \mu_t + \sigma_t \epsilon_t \text{ for } t = 1, 2, \cdots, \tag{1.1}$$

where μ_t is the conditional mean, σ_t is the square root of the conditional variance (volatility) and $\{\epsilon_t\}$ is an independent, identically distributed process with mean zero, variance one and marginal distribution F_ϵ. Based on a sample $\{y_t\}_{t=1}^n$, the first stage of the estimation produces $\hat{\mu}_t$ and $\hat{\sigma}_t$. In the second stage, $e_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}$ for $t = 1, ..., n$ are used to estimate a generalized pareto density approximation for the tails of f_t, which in turn produce VaR

[1] We will assume throughout this paper that F_t is absolutely continuous.

and TailVaR sequences for the conditional distribution of Y_t. Backtesting of their method (using various financial return series) against some widely used *direct* estimation methods that assume a specific form for the distribution of ϵ_t (gaussian, student-t) have produced favorable, albeit specific results. Although encouraging, results from backtesting could be specific to the series and periods analyzed, providing limited information regarding the statistical properties of the two stage estimators for VaR and TailVaR. Furthermore, since various first and second stage estimators could be proposed, the question of how to best implement the two stage estimators remains unexplored.

In this paper we make three contributions to the growing literature on VaR and TailVaR estimation. First, we propose a nonparametric markov chain model (Härdle and Tsybakov, 1997 and Hafner, 1998) for $\{Y_t\}$ dynamics, as well as an improved nonparametric estimation procedure for the conditional volatility of the returns (Fan and Yao, 1998) in the first stage of estimation. The objective is to have in place a model that is general enough to accommodate nonlinearities that have been regularly verified in empirical work (Andreou et al., 2001, Hafner, 1998, Patton, 2001 and Tauchen, 2001). If the assumptions on the (parametric) structure of μ_t, σ_t and ϵ_t are not sufficiently general, there is a distinct possibility that the resulting sequence of residuals will be inadequate for the EVT based approach that follows in the second stage. This might be particularly true for popular parametric models of conditional mean and volatility of financial returns (GARCH, ARCH and their many relatives), specifically with regards to volatility asymmetry.

Second, we propose an alternative estimation procedure for the EVT inspired parametric tail model in the second stage based on L-Moment Theory (Hosking, 1990). The L-Moment estimators we use are easier and faster to implement, and in finite samples outperform the constrained maximum likelihood estimation methods that have prevailed in the empirical finance literature. Since L-Moment estimation is not commonly used in econometrics or empirical finance, we provide a brief introduction and discussion in section 2.

The third contribution we make to this literature is in the form of a Monte Carlo study. As noted previously, the statistical properties of the VaR and TailVaR estimators that result from the two stage procedure discussed above are unknown both in finite samples and asymptotically. In addition, since it is possible to envision various alternative estimators for the first and second stages of the procedure, a number of final estimators of VaR and TailVaR emerge. To assess their properties as well as to shed some light on their relative performance we design an extensive Monte Carlo study. Our study considers various data generating processes (DGPs) that mimic the empirical regularities of financial time series, including asymmetric conditional volatility, leptokurdicity, infinite past memory and asymmetry of conditional return distributions. The ultimate goal here is to provide empirical researchers with some guidance on how to choose between a set of VaR and TailVaR estimators. Our simulations indicate that the estimation strategy we propose outperforms, as measured by the estimators' mean squared error, the method proposed by McNeil and Frey. We also backtest our estimation procedure on four financial time series (Dow 30 Industrial Stock Price Index, Microsoft stock, Nasdaq and the S&P 500) covering different time periods. Besides this introduction, the paper has five additional sections. Section 2 discusses the stochastic model and proposed estimation. Section 3 describes the Monte Carlo design, section 4 summarizes the results and section 5 contains a backtesting evaluation of the estimators considered. Section 6 provides a brief conclusion.

2 Stochastic Properties of $\{Y_t\}$ and Estimation

Estimation of discrete time stochastic processes such as (1.1) to model asset price returns has proceeded by making specific assumptions on μ_t, σ_t and ϵ_t. It is commonly assumed that the conditional mean μ_t and the conditional volatility σ_t have known parametric structure and that the conditioning set depends only on past realizations of the process.[2] In addition, to facilitate estimation, specific distributional assumptions are normally made on ϵ_t. Estimation of models such as ARCH, GARCH, EGARCH, IGARCH and many other derived parametric variants follows this general description. The adequacy of these models in fitting observed data, producing accurate forecasts, and the ease with which they can be estimated depends largely on these assumptions. In fact, the great profusion of ARCH type models is the result of an attempt to accommodate empirical regularities that have been repeatedly observed in financial return series. More recently, a number of papers have proposed nonparametric and/or semiparametric modeling (Carroll, Härdle and Mammen, 2002, Härdle and Tsybakov, 1997, Masry and Tjøstheim, 1995) of μ_t, σ_t, as well as less restrictive assumptions on ϵ_t. These flexible nonparametric models for $\{Y_t\}$ are more difficult to estimate than their parametric counterparts, but there can be substantial inferential gains if the alternative parametric models are misspecified or unduly restrictive.

Our interest is in obtaining estimates for α-VaR and α-TailVaR associated with the conditional density f_t, where in general conditioning is on the filtration $M_{t-1} = \sigma(\{Y_s : M < s \leq t-1\})$, where $-\infty \leq M \leq t-1$. We denote such conditional densities by $f(Y_t|M_{t-1})$ for $t = 2, 3, \ldots$. Letting $q(\alpha) = F_\epsilon^{-1}(\alpha)$ be the quantile of F_ϵ and given that $F_\epsilon(x) = F(\mu_t + \sigma_t x|M_{t-1})$ we have that α-VaR for $f(y|M_{t-1})$ is given by,

$$F^{-1}(\alpha|M_{t-1}) = \mu_t + \sigma_t q(\alpha). \tag{2.1}$$

Similarly, α-TailVaR for $f(y|M_{t-1})$ is given by,

$$E\left(Y_t|Y_t > F^{-1}(\alpha|M_{t-1}), M_{t-1}\right) = \mu_t + \sigma_t E(\epsilon_t|\epsilon_t > q(\alpha)). \tag{2.2}$$

Hence, the estimation of α-VaR and α-TailVaR can be viewed as a process of estimation for the unknown functionals in (2.1) and (2.2). We start by considering the following nonparametric specifications for μ_t, σ_t and the process Y_t. Assume that $\{(Y_t, Y_{t-1})'\}$ is a two dimensional strictly stationary process

[2] But see Shepherd (1996) for alternative modeling strategies.

with conditional mean function $E(Y_t|Y_{t-1} = x) = m(x)$ and conditional variance $E\left((Y_t - m(x))^2|Y_{t-1} = x\right) = \sigma^2(x) > 0$. The process is described by the following markov chain of order 1,

$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\epsilon_t \text{ for } t = 1, 2, \dots , \qquad (2.3)$$

where ϵ_t is an independent strictly stationary process with unknown marginal distribution F_ϵ that is absolutely continuous with mean zero and unit variance. Note that the conditional skewness, $\alpha_3(x)$ and kurtosis, $\alpha_4(x)$ of the conditional density of Y_t given $Y_{t-1} = x$ are given by, $\alpha_3(x) = E(\epsilon_t^3)$ and $\alpha_4(x) = E(\epsilon_t^4)$. We assume that such moments exist and are continuous and that $m(x)$ and $\sigma^2(x)$ have uniformly continuous second order derivatives on an open set containing x.

Recursion (2.3) is the conditional heterocedastic autoregressive nonlinear (CHARN) model of Diebolt and Guègan (1993), Härdle and Tsybakov (1997) and Hafner (1998). It is a special case (one lag) of the nonlinear-ARCH model treated by Masry and Tjøstheim (1995). The CHARN model provides a generalization for the popular GARCH(1,1) model in that $m(x)$ is a nonparametric function, and most importantly $\sigma^2(x)$ is not a linear function of Y_{t-1}^2. The symmetry in Y_{t-1} of the conditional variance in GARCH models is a particularly undesirable restriction when modeling financial time series due to the empirically well documented *leverage effect* (Chen, 2001, Ding et al., 1993, Hafner, 1998 and Patton, 2001). However, (2.3) is more restrictive than traditional GARCH models in that its markov property restricts its ability to effectively model the longer memory that is commonly observed in return processes.[3] Estimation of the CHARN model is relatively simple and provides much of its appeal in our context.

2.1 First Stage Estimation

The estimation of $m(x)$ and $\sigma^2(x)$ in (2.3) was considered by Härdle and Tsybakov (1997). Unfortunately, their procedure for estimating the conditional variance $\sigma^2(x)$ suffers from significant bias and does not produce estimators that are constrained to be positive. Furthermore, the estimator is not asymptotically design adaptive to the estimation of m, i.e., the asymptotic properties of their estimator for conditional volatility is sensitive to how well m is estimated. We therefore consider an alternative estimation procedure due to Fan and Yao (1998), which is described as follows. First, we estimate $m(x)$ using the local linear estimator of Fan (1992). Let $W(x), K(x) : \Re \to \Re$ be symmetric kernel functions, y in the support of the conditional density of Y_t and

[3]The model of Masry and Tjøstheim (1995) and equations (2) and (3) in Carrol, Härdle and Mammen (2002) provide a full nonparametric generalization of ARCH and GARCH(1,1) models. However, nonparametric estimators for the latter model are unavailable and for the former, convergence of the proposed estimators for m and σ^2 is extremely slow as the number of lags in the conditioning set increases (curse of dimensionality).

$h(n)$, $h_1(n)$ be sequences of positive real numbers - bandwidths - such that $h(n), h_1(n) \to 0$ as $n \to \infty$. Let

$$(\hat{\zeta}, \hat{\zeta}_1)' = argmin_{\zeta,\zeta_1} \sum_{t=2}^{n} (y_t - \zeta - \zeta_1(y_{t-1} - y))^2 K\left(\frac{y_{t-1} - y}{h(n)}\right),$$

then the local linear estimator of $m(y)$ is $\hat{m}(y) = \hat{\zeta}(y)$. Second, let $\hat{r}_t = (y_t - \hat{m}(y_{t-1}))^2$, and define,

$$(\hat{\eta}, \hat{\eta}_1)' = argmin_{\eta,\eta_1} \sum_{t=2}^{n} (\hat{r}_t - \eta - \eta_1(y_{t-1} - y))^2 W\left(\frac{y_{t-1} - y}{h_1(n)}\right),$$

then the local linear estimator of $\sigma^2(y)$ is $\hat{\sigma}^2(y) = \hat{\eta}(y)$.

It is clear that an important element of the nonparametric estimation of m and σ^2 is the selection of the sequence of bandwidths $h(n)$ and $h_1(n)$. We select the bandwidths using the data driven plug-in method of Ruppert et al. (1995) and denote them by $\hat{h}(n)$ and $\hat{h}_1(n)$. $\hat{h}(n)$ and $\hat{h}_1(n)$ are obtained based on the following regressand-regressor sequences $\{(y_t, y_{t-1})\}_{t=2}^n$ and $\{(\hat{r}_t, y_{t-1})\}_{t=2}^n$, respectively. This bandwidth selection method is theoretically superior to the popular cross-validation method and is a consistent estimator of the (optimal) bandwidth sequence that minimizes the asymptotic mean integrated squared error of $\hat{m}$ and $\hat{\sigma}^2$.[4] We chose a common kernel function (gaussian) in implementing our estimators. In the context of the CHARN model the first stage estimators for μ_t and σ_t^2 in (2.1) and (2.2) are respectively $\hat{m}(y_{t-1})$ and $\hat{\sigma}^2(y_{t-1})$.

2.2 Second Stage Estimation

In the second stage of the estimation we obtain estimators for $q(\alpha)$ and $E(\epsilon_t|\epsilon_t > q(\alpha))$. The estimation is based on a fundamental result from extreme value theory, which states that the distribution of the exceedances of any random variable (ϵ) over a specified nonstochastic threshhold u, i.e, $Z = \epsilon - u$ can be suitably approximated by a generalized pareto distribution - GPD (with location parameter equal to zero) given by,

$$G(x; \beta, \psi) = 1 - \left(1 + \psi \frac{x}{\beta}\right)^{-1/\psi}, x \in D \tag{2.4}$$

where $D = [0, \infty)$ if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$.[5]

[4] See Ruppert et al. (1995) and Fan and Yao (1998). For an alternative estimator of $\sigma^2(x)$ see Ziegelmann (2002).

[5] See Pickands (1975) and Embrechts et al. (1997) for a complete characterization of the result.

First stage estimators $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ can be used to produce a sequence of standardized residuals $\left\{e_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}\right\}_{t=1}^{n}$ which can be used to estimate the tails of f_ϵ based on (2.4). For this purpose we order the residuals such that $e_{j:n}$ is the j^{th} largest residual, i.e., $e_{1:n} \geq e_{2:n} \geq ... \geq e_{n:n}$ and obtain $k < n$ excesses over $e_{k+1:n}$ given by $\{e_{j:n} - e_{k+1:n}\}_{j=1}^{k}$, which will be used for estimation of a GPD. By fixing k we in effect determine the residuals that are used for tail estimation and randomly select the threshold. It is easy to show that for $\alpha > 1 - k/n$ and estimates $\hat{\beta}$ and $\hat{\psi}$, $q(\alpha)$ and $E(\epsilon_t|\epsilon_t > q(\alpha))$ can be estimated by,

$$\widehat{q(\alpha)} = e_{k+1:n} + \frac{\hat{\beta}}{\hat{\psi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\psi}} - 1\right) \tag{2.5}$$

and for $\psi < 1$

$$\hat{E}(\epsilon_t|\epsilon_t > q(\alpha)) = \widehat{q(\alpha)}\left(\frac{1}{1-\hat{\psi}} + \frac{\hat{\beta} - \hat{\psi}e_{k+1:n}}{(1-\hat{\psi})\widehat{q(\alpha)}}\right). \tag{2.6}$$

It is clear that these estimators and their properties depend on the choice of k. This question has been studied by McNeil and Frey (2000) and is also addressed in the Monte Carlo study in section 4 of this paper. Combining the estimators in (2.5) and (2.6) with first stage estimators, and using (2.1) and (2.2) gives estimators for $\alpha - VaR$ and $\alpha - TailVaR$.

2.3 L-Moment Estimation of β and ψ

Given the results in Smith (1984, 1987), estimation of the GPD parameters has normally been done by constrained maximum likelihood (ML). Here we propose an alternative estimator based on L-Moment Theory (Hosking, 1990 and 1997). Traditionally, raw moments have been used to describe the location, scale, and shape of distribution functions. L-Moment Theory provides an alternative approach that exhibits a number of desirable properties. Here, we provide a brief summary and justification for its use and direct the reader to Hosking's papers for a thorough coverage and understanding.

Let F_ϵ be a distribution function associated with a random variable ϵ and $q(u) : (0,1) \to \Re$ its quantile. The r^{th} L-moment of ϵ is defined as,

$$\lambda_r = \int_0^1 q(u)P_{r-1}(u)du \text{ for } r = 1, 2, ... \tag{2.7}$$

where $P_r(u) = \sum_{k=0}^{r} p_{r,k}u^k$ and $p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$, which contrasts with the traditional raw moments $\mu_r = \int_0^1 q(u)^r du$. Theorem 1 in Hosking (1990) gives the following justification for using L-moments to describe distributions: a) μ_1 is finite if and only if λ_r exist for all r; b) a distribution F_ϵ with finite μ_1 is uniquely characterized by λ_r for all r. Thus, a distribution can be characterized by its L-moments even if raw moments of order greater than 1 do not exist, and most importantly, this characterization is unique, which is not true for raw moments.

It is easily verified that $\lambda_1 = \mu_1$, therefore the first L-moment when it exists provides the traditionally used measure of location for a distribution. As pointed out by Hosking (1990 and 1997), λ_2 is up to a scalar the expectation of Gini's mean difference statistic, therefore providing a measure of scale that differs from the traditional variance - $\mu_2 - \mu_1^2$ by placing smaller weights on differences between realizations of the random variable. Hosking (1989) shows that if μ_1 exists $-1 < \tau_3 \equiv \frac{\lambda_3}{\lambda_2} < 1$ with $\tau_3 = 0$ for symmetric distributions, providing a bounded measure of skewness that is less sensitive to the extreme tails of the distribution than the traditional (unbounded) measure of skewness given by $\frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}$. Similarly, $-1 < \tau_4 \equiv \frac{\lambda_4}{\lambda_2} < 1$ can be interpreted as a bounded measure of kurtosis (see Oja, 1981) that is less sensitive to the extreme tails of the distribution than the traditional (unbounded) measure given by $\frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2}$. Hence, contrary to traditional measures of location and shape, L-moment based measures of scale, skewness and kurtosis do not require the existence of higher order raw moments, allowing for synthetic measures of distribution shape even when higher order raw moments do not exist. The use of these alternative measures of shape can be particularly useful in empirical finance and in financial risk management in particular.

In addition, L-moments can be used to estimate a finite number of parameters $\theta \in \Theta$ that identify a member of a family of distributions. Suppose $\{F_\epsilon(\theta) : \theta \in \Theta \subset \Re^p\}$, p a natural number, is a family of distributions which is known up to θ. A sample $\{\epsilon_t\}_{t=1}^T$ is available and the objective is to estimate θ. Since, λ_r, $r = 1, 2, 3...$ uniquely characterizes F, θ may be expressed as a function of λ_r. Hence, if estimators $\hat{\lambda}_r$ are available, we may obtain $\hat{\theta}(\hat{\lambda}_1, \hat{\lambda}_2, ...)$. From (2.7), $\lambda_{r+1} = \sum_{k=0}^{r} p_{r,k}\beta_k$ where $\beta_k = \int_0^1 q(u)u^k du$ for $r = 0, 1, 2, \cdots$. Given the sample, we define $\epsilon_{k,T}$ to be the k^{th} smallest element of the sample, such that $\epsilon_{1,T} \leq \epsilon_{2,T} \leq ... \leq \epsilon_{T,T}$. An unbiased estimator of β_k is

$$\hat{\beta}_k = T^{-1} \sum_{j=k+1}^{T} \frac{(j-1)(j-2)...(j-k)}{(T-1)(T-2)...(T-k)} \epsilon_{j,T}$$

and we define $\hat{\lambda}_{r+1} = \sum_{k=0}^{r} p_{r,k}\hat{\beta}_k$ for $r = 0, 1, \cdots, T-1$.

In particular, if F_ϵ is a generalized pareto distribution with $\theta = (\mu, \beta, \psi)$, it can be shown that $\mu = \lambda_1 - (2-\psi)\lambda_2$, $\beta = (1-\psi)(2-\psi)\lambda_2$, $\psi = -\frac{1-3(\lambda_3/\lambda_2)}{1+(\lambda_3/\lambda_2)}$. In our case, where $\mu = 0$, $\beta = (1-\psi)\lambda_1$, $\psi = 2 - \lambda_1/\lambda_2$ we define the following L-moment estimators for ψ and β,

$$\hat{\psi} = 2 - \frac{\hat{\lambda}_1}{\hat{\lambda}_2} \text{ and } \hat{\beta} = (1-\hat{\psi})\hat{\lambda}_1.$$

Similar to ML estimators, these L-moment parameter estimators are $\sqrt{T}$-asymptotically normal for $\psi < 0.5$. However, they are much easier to compute than ML estimators as no numerical optimization or iterative procedure is necessary. Although asymptotically inefficient relative to ML estimators, L-moment based parameter estimators have reasonably high asymptotic efficiency (see Hosking, 1990). For the GPD considered here, asymptoic efficiency is always higher than 70 percent when $0 < \psi < 0.3$. Similar levels of asymptotic efficiency can also be verified in the estimation of Generalized Extreme Value - GEV (Hosking et al., 1985), Gumbel and Normal distributions.

More important, from a practical perspective, is that L-Moment based parameter estimators can outperform ML (based on mean squared error) in finite samples as indicated by Hosking et al. (1985), Hosking (1987) and the Monte Carlo we conduct in section 3. The results are not entirely surprising as the efficiency of ML estimators is attained only asymptotically. In fact, as observed by Hosking and Wallis (1987), it may be necessary to deal with very large samples before asymptotic distributions provide useful approximations to their finite sample equivalents. This seems to be especially true for GPD and GEV estimation, but it can also be verified in other more general contexts.[6]

2.4 Alternative First Stage Procedures

Here we define two alternatives to the first stage estimation discussed above. The first is the quasi maximum likelihood estimation (QMLE) method used by McNeil and Frey. In essence it involves estimating by maximum likelihood the following regression model,

$$Y_t = \theta_1 Y_{t-1} + \sigma_t \epsilon_t \text{ for } t = 1, 2, ... \tag{2.8}$$

where $\epsilon_t \sim NIID(0,1)$ and $\sigma_t^2 = \gamma_0 + \gamma_1(Y_{t-1} - \theta_1 Y_{t-2})^2 + \gamma_2 \sigma_{t-1}^2$. We will refer to this procedure as GARCH-N (Bollerslev, 1986). The second alternative estimator we consider is identical to the first procedure but assumes that the ϵ_t are iid with a standardized Student-t density, denoted by $f_s(\nu)$, where $\nu > 2$ is a parameter to be estimated together $\theta_1, \gamma_0, \gamma_1, \gamma_2$ by maximum likelihood. This estimator has gained popularity in that the Y_t inherits the leptokurdicity of ϵ_t, a characteristic of financial asset returns that have been abundantly reported in the literature. We refer to this procedure as GARCH-T.

Obviously, a number of other first stage estimators can be considered.[7] Our choice of alternative estimators to be considered in the Monte Carlo study that follows was mostly guided not by the desire to be exhaustive, but rather an attempt to represent what is commonly used both in the empirical finance literature and in practice.

[6]See, e.g., Hannan (1987) and Mandy and Martins-Filho (2001).

[7]For a list of many alternatives see Gourieroux (1997).

3 Monte Carlo Design

In this section we describe and justify the design of the Monte Carlo study. Our study has two main objectives. First, to provide evidence on the finite sample distribution of the various estimators proposed in the previous section and second, to evaluate the relative performance of the estimators. This is, to our knowledge, the first evidence on the finite sample performance of the two-stage estimation procedures described above. Second, to provide applied researchers with some guidance over which estimators to use when estimating VaR and TailVaR.

In designing our Monte Carlo experiments we had two goals. First, the data generating process (DGP) had to be flexible enough to capture the salient characteristics of time series on asset returns. Second, to reduce specificity problems, we had to investigate the behavior of the estimators over a number of relevant values and specifications of the design parameter and functions in the Monte Carlo DGP.

3.1 The base DGP

The main DGP we consider is a nonparametric GARCH model first proposed by Hafner (1998) and later studied by Carroll, Härdle and Mammen (2002). We assume that $\{Y_t\}$ is a stochastic process representing the log-returns on a financial asset with $E(Y_t|M_{t-1}) = 0$ and $E(Y_t^2|M_{t-1}) = \sigma_t^2$, where $M_{t-1} = \sigma(\{Y_s : M < s \leq t-1\})$, where $-\infty \leq M \leq t-1$. We assume that the process evolves as,

$$Y_t = \sigma_t \epsilon_t \text{ for } t = 1, 2, ... \tag{3.1}$$

$$\sigma_t^2 = g(Y_{t-1}) + \gamma \sigma_{t-1}^2 \tag{3.2}$$

where $g(x)$ is a positive, twice continuously differentiable function and $0 < \gamma < 1$ is a parameter. $\{\epsilon_t\}$ is assumed to be a sequence of independent and identically distributed random variables with the skewed Student-t density function. This density was proposed by Hansen (1994) and is given by

$$f(x; v, \lambda) = \begin{cases} bc\left(1 + \frac{1}{v-2}\left(\frac{bx+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} & \text{for } x \geq -a/b \\ bc\left(1 + \frac{1}{v-2}\left(\frac{bx+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} & \text{for } x \leq -a/b \end{cases}$$

where $c \equiv \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi(\nu-2)}}$, $a \equiv 4\lambda c\left(\frac{\nu-2}{\nu-1}\right)$, $b \equiv \sqrt{1 + 3\lambda^2 - a^2}$. Hansen proved that $E(\epsilon_t) = 0$ and $V(\epsilon_t) = 1$. The following lemma gives expressions for the skewness and kurtosis of the asymmetric Student-t density.

Lemma 1: *Let $f(x; v, \lambda)$ be the skewed t-Student density function of Hansen (1994). Let κ_i for $i = 1, 2, 3, 4$ be as defined in* **Proposition 1** *in the appendix,* *then the skewness* α_3 *of the density is given by,*

$$\alpha_3 = \frac{8\kappa_3}{b^3}(\lambda^3+\lambda) - \frac{6\kappa_2 a}{b^3}(\lambda^3+3\lambda) + \frac{12a^2\kappa_1}{b^3}\lambda - \frac{a^3+3a(1-\lambda)^3}{b^3},$$

and its kurtosis is given by,

$$\begin{aligned}\alpha_4 &= \frac{\kappa_4}{b^4}(2\lambda^5+20\lambda^3+10\lambda) - \frac{32a\kappa_3}{b^4}(\lambda^3+\lambda) + \frac{12a^2\kappa_2}{b^4}(\lambda^3+3\lambda) - \frac{16a^3\lambda\kappa_1}{b^4} \\ &+ \frac{1}{b^4}\left(a^4 + \frac{3(1-\lambda)^5(v-2)}{v-4} + 6a^2(1-\lambda)^3\right).\end{aligned}$$

It is clear from these expressions that skewness and kurtosis are controlled by the parameters λ and ν. When $\lambda = 0$ the distribution is a symmetric standardized Student-t. The α-VaR for ϵ_t was obtained by Patton (2001) and is given by,

$$\alpha - VaR = \begin{cases} \frac{1-\lambda}{b}\sqrt{\frac{\nu-2}{\nu}}F_s^{-1}\left(\frac{\alpha}{1-\lambda};\nu\right) - \frac{a}{b} & \text{for } 0 < \alpha < \frac{1-\lambda}{2} \\ \frac{1+\lambda}{b}\sqrt{\frac{\nu-2}{\nu}}F_s^{-1}\left(0.5 + \frac{1}{1+\lambda}\left(\alpha - \frac{1-\lambda}{2}\right);\nu\right) - \frac{a}{b} & \text{for } \frac{1-\lambda}{2} \le \alpha < 1 \end{cases},$$

where F_s is the cumulative distribution of a random variable with Student-t density and ν degrees of freedom. In the following lemma we obtain α-TailVaR for ϵ_t when $\alpha \geq -\frac{a}{b}$.

Lemma 2: *Let* X *be a random variable with density function given by an asymmetric Student-t and define the truncated density,*

$$f_{X>z}(x; v, \lambda) = \frac{f(x; v, \lambda)}{1 - F(z)} \text{ for } z \geq -a/b$$

where F *is the distribution function of* X. *Then, the expected shortfall of* X, $E(X|X>z) = \int_z^\infty x f_{X>z}(x;v,\lambda)dx$ *is given by,*

$$\begin{aligned}E(X|X>z) &= (1-F(z))^{-1}\left(\frac{c(1+\lambda)^2}{b}\left(\frac{v-2}{v-1}\right)\beta^{-(v-1)/2}\right. \\ &\left. - \frac{(1+\lambda)a}{b}\left(1 - F_s\left(\frac{bz+a}{1+\lambda}\sqrt{\frac{v}{v-2}}\right)\right)\right)\end{aligned}$$

where $\beta = \left(cos\left(arctan\left(\frac{bz+a}{(1+\lambda)\sqrt{v-2}}\right)\right)\right)^2$ *and* F_s *is the cumulative distribution of a random variable with Student-t density and* v *degrees of freedom.*

Under (3.1), (3.2) and the assumptions on ϵ_t, it is easy to verify that the conditional skewness $\alpha_3(Y_t|M_{t-1}) = E(\epsilon_t^3)$ and the conditional kurtosis $\alpha_4(Y_t|M_{t-1}) = E(\epsilon_t^4)$.

This DGP incorporates many of the empirically verified regularities normally ascribed to returns on financial assets: (1) asymmetric conditional variance with higher volatility for large negative returns and smaller volatility for positive returns (Hafner, 1998); (2) conditional skewness (Aït-Sahalia and Brandt, 2001, Chen, 2001, Patton, 2001); (3) Leptokurdicity (Tauchen, 2001, Andreou et al., 2001); and (4) nonlinear temporal dependence. Our objective, of course, was to provide a DGP design that is flexible enough to accommodate these empirical regularities and to mimic the properties observed in return time series.

We designed a total of 144 experiments over the base DGP described above. Table 1 provides a numbering scheme for the experiments that is used in the description of the Monte Carlo results. In summary, there are two sample sizes considered for the first stage of the estimation $n_S = \{500, 1000\}$, three values for γ, $n_\gamma = \{0.3, 0.6, 0.9\}$, three values for λ, $n_\lambda = \{0, -0.25, -0.5\}$, two values for α, $n_\alpha = \{0.95, 0.99\}$, two values for the number of observations used in the second stage of the estimation k, $n_k = \{60, 100\}$, and two functional forms for $g(x)$, which we denote by $g_1(x) = 0.5 + \frac{exp(-4x)}{1+exp(-4x)}$ and $g_2(x) = 1 - 0.9exp(-2x^2)$. The total number of replications in the Monte Carlos is held fixed at 1000 for all experiments. Graphs 1A and 1B in Appendix 2 provide the general shape for these volatility specifications together with that implied by GARCH type models. We now turn to the results of our Monte Carlo.

4 Results

We considered a total of seven estimators for VaR and TailVaR. There are three first stage estimators: the nonparametric method we propose, GARCH-N, GARCH-T; and two second stage estimators: the L-moments based estimator we propose and the ML estimator. In addition we consider a direct estimation method that estimates VaR and TailVaR using our nonparametric method in the first stage and assuming the estimated ϵ_t in (3.1) is indeed distributed as an asymmetric Student-t density. In this case, all parameters are estimated *via* maximum likelihood. Since this direct ML estimator is based on a correct specification of the conditional density we expect it to outperform all other methods provided that the nonparametric residuals from the first stage are suitable estimates for ϵ_t. Our focus is on the remaining estimators, which are all based on stochastic models that are misspecified relative to the DGPs considered. Specifically, the nonparametric estimator is based on a model in which the volatility function is assumed to depend only on Y_{t-1} rather than the entire history of time series (markov property of order 1). The GARCH type models are misspecified in that g in our DGPs is not linear in Y_{t-1}. We implement our nonparametric local linear estimator using a Gaussian kernel and a theoretical optimal bandwidth. A summary of simulation results is provided in Appendix 2.

General Results on Relative Performance: Tables 2A-2F provide the ratio between an estimator's mean squared error (MSE) and the MSE for the direct estimation method, which we call relative MSE, for $n = 1000$. As expected, in virtually all experiments, relative MSE> 1 indicating that the direct estimation method (correctly specified DGP) outperforms all other estimators for VaR and TailVaR. Exceptions occur almost exclusively in experiments with $\gamma = 0.9$, which is expected, given that for these experiments the nonparameteric residuals from the first stage are most likely to be poor estimates for the true errors due to the strong memory of the volatility process. Interestingly, there are a series of very general conclusions that can be reached regarding the relative performance of the other estimators:[8] 1) for both VaR and TailVaR estimation and all estimators considered, second stage estimation based on L-moments produces smaller MSE than when based on ML. This conclusion holds for most experiments, volatilities and λ. The performance of ML based estimators seems to improve when $k = 100$, n and in estimating TailVaR, but not significantly. The improvement of ML based estimators when $k = 100$ is consistent with what one one expect from asymptotic theory and also confirms the results in Hosking and Wallis (1987). 2) VaR and TailVaR estimators based on the nonparametric method produce lower MSE for virtually all experimental designs with $\gamma = 0.3, 0.6$. Estimators based on GARCH-T are consistently the second best option. For experiments with $\gamma = 0.9$, GARCH-T estimator performs better, which is to be expected since in this case the true volatility in the DGP deviates most from markov property of order 1 assumed in the nonparameteric model. We detect no decisive impact of λ, n and other design parameters on the relative performance of the estimators.

These results generally indicate that the combined nonparametric/L - moment estimation procedure we propose is superior to GARCH/ML type estimators in virtually all experimental designs with low to moderate γ parameters. Furthermore, since our estimator assumes that $\{Y_t\}$ is markov of order 1, contrary to GARCH type models, our results reveal that except in the most extreme case, where $\gamma = 0.9$ nonlinearities in volatility may be more important to VaR and TailVaR estimation performance than accounting for the non-markov property of the series. In fact, given that $g_1(x)$ produces conditional volatilities that are similar to those empirically obtained in Hafner (1998), it seems warranted to conclude that this would indeed be the case for some typical time series of asset returns. Support for this conjecture is given on section 5 where a backtesting evaluation of the estimators using some common time series of asset returns is performed.

We also found that in finite samples, VaR and TailVaR estimators based on more sophisticated nonparametric estimators in the first stage, such as that proposed by Carroll, Härdle and Mammen (2002), did not outperform our nonparametric procedure in finite samples. To provide some numerical evidence, we implement the first stage volatility estimator proposed by Carroll et al. (2002) with $J = 2$, using their proposed method based on least squares, employing the *Rule-of-Thumb* bandwidth selection for the bivariate marginal integration estimation as in Linton and Nielson (1995). In the second stage we use the estimated residual to calculate the VaR and TailVaR using L-Moments. We compare the performance of their estimator and our nonparametric/L-

[8] Unless explicitly noted conclusions are also valid for the cases when $n = 500$.

Moments estimator based on MSE for a subset of the 144 experiments. Here we report only the representative results that emerge from the fifth experimental design with $\lambda = -0.25$ and $n = 500$. When the volatility function is $g_1(x)$, the MSE for out method $MSE_{NP} = 0.227$ and for their method $MSE_C = 0.661$ in estimating VaR, $MSE_{NP} = 0.865$ and $MSE_C = 1.793$ in estimating TailVaR. When the volatility function is $g_2(x)$, $MSE_{NP} = 0.196$ and $MSE_C = 0.890$ in estimating VaR, $MSE_{NP} = 0.457$ and $MSE_C = 3.769$ in estimating TailVaR. The results in terms of the relative performances are not changed when Carroll et al.'s method based on errors in variables/minimum distance in the first stage or ML in the second stage is employed.

Having derived some general conclusions regarding the relative performance of the various estimators, we now turn to a discussion of their bias, MSE and how these measures are impacted by the experimental designs.

Results on MSE: Tables 3A-F show MSE for VaR and TailVaR based on the different estimators considered. Results reported in the tables are based on using L-Moments in the second stage. Unless explicitly mentioned, the conclusions are also valid for the cases where ML is used in the second stage. We first examine the impact of the sample size n. The MSE for the nonparametric method falls with increased n for virtually all design parameters and volatilities. There is weak evidence that this regularity holds for GARCH type estimators, with violations showing up more frequently for GDP based on g_2, see tables 3D-F.

The number of observations used in the second stage - $k-$has no distinguishable impact on the MSE for any of the estimators considered holding all other design parameters fixed. This is verified by comparing rows 1 and 2, 3 and 4, and so on until rows 11 and 12 in tables 3A-F. Insensitivity of the GARCH-N/ML VaR estimator to changes in k in this range was also obtained in a simulation study by McNeil and Frey (2000). This is most likely due to the range of k we have used.

Ceteris Paribus, the estimators' MSE increases for VaR and TailVaR when the quantile increses from 0.95 to 0.99 across all design parameters and volatilities. This can be verified by comparing rows 1 and 3, 2 and 4, and so on until rows 10 and 12 in tables 3A-F. Thus, it seems that estimation of VaR and TailVaR associated with larger quantiles is more difficult for all estimation methods.

As expected, the MSE for the nonparametric estimator of VaR and TailVaR falls with the value of γ. This can be verified by comparing rows 1, 5 and 9; 2, 6 and 10, and so on until 4, 8 and 12 in tables 3A-F. However, the impact of γ on MSE of GARCH type estimators is ambiguous for DGP based on g_2 across different design parameters, with weak evidence that GARCH type estimators perform better for DGP based on g_1 with smaller value of γ.

The MSE for all estimators of VaR and TailVaR estimators decrease significantly with λ across all parameter designs and volatilities, specially when considering the nonparametric procedure. This sensitivity of the MSE to λ is most likely explained by the fact that in our DGPs $\lambda \leq 0$ and the data is skewed towards the positive quadrant. As such, in the second stage estimation we are selecting data that are larger than the k^{th} order statistic and considering the 95 and 99 percent quantiles. Hence, more representative data of tail behavior is used when λ decreases. Finally, the MSE across all design parameters, volatilities and estimators is significantly larger when estimating TailVaR compared to VaR. Hence, our simulations seem to indicate that the estimation of TailVaR is more difficult than VaR, at least as measured by MSE. The result is largely due to increased variance in TailVaR estimation rather than bias, since as indicated below, there is no clear pattern for the change in the bias.

Results on Bias: The impact of various design parameters on the bias in VaR and TailVaR estimation across the procedures, design parameters and volatilities is much less clear and definitive than that on MSE.[9] In particular, as the sample size n or k increases there is no clear impact on bias. Hence, the reductions on MSE with increased n that are reported above are largely due to a reduction on variance. The most definite result regarding estimation bias we could infer from the simulations is that all estimation procedures seem to have a positive bias in the estimation of both VaR and TailVaR. We illustrate this fact by plotting the bias for the two best estimators (Nonparametric and GARCH-T) for $\lambda = -0.25$ and $n = 1000$ across all experimental designs, which are represented in the horizontal axis. Graphs 2A and 2C illustrate the bias in estimating VaR for g_1 and g_2, respectively, and we note that virtually all biases are positive. Graphs 2B and 2D illustrate the bias in estimating TailVaR for g_1 and g_2, respectively, and once again we observe that virtually all biases are positive.[10] The exceptions occur exclusively for experimental designs that have $\gamma = 0.9$. The other regularity that we observe is that in most DGPs the nonparametric method seems to have smaller bias than the GARCH type estimator. Once again, violations occur when γ is large.

5 Evaluation via Backtesting

In this section we backtest the estimators considered in the previous section using four historical daily series on the following price indices: (1) Dow 30 Industrial Stock Price Index from November 1996 to March 2000; (2) Microsoft Corporation common stock price from May 1998 to May 2002; (3) Nasdaq composite index from February 1981 to January 1985, and the (4) Standard and Poor's 500 composite stock price index from January 1965 to January 1969. The data are obtained from website *http://www.economy.com* and *http://finance.yahoo.com*. The return for each series is calculated as $y_t = 100 \times \ln(\frac{P_t}{P_{t-1}})$, where P_t refers to the price level at time t.

To perform the backtesting of the estimators on a data set $\{y_1, y_2, \cdots, y_m\}$, we utilize the previous n observations $\{y_{t-n+1}, y_{t-n+2}, \cdots, y_t\}$ to estimate the $\alpha-$VaR by $\hat{F}^{-1}(\alpha|M_t)$ and the $\alpha-$TailVaR by $\hat{E}(Y_{t+1}|Y_{t+1} > F^{-1}(\alpha|M_t), M_t)$ for $n_\alpha = \{0.95, 0.99\}$, where $0 < n < m$, $t \in T = \{n, n+1, \cdots, m-1\}$,

[9] Tables similar to 3A-F for the bias are available upon request.

[10] Similar graphs result from $\lambda = 0, -0.5$ and $n = 500$

$\hat{F}^{-1}(\alpha|M_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{q}(\alpha)$ and $\hat{E}(Y_{t+1}|Y_{t+1} > F^{-1}(\alpha|M_t), M_t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{E}(\epsilon_{t+1}|\epsilon_{t+1} > q(\alpha))$. Henceforth, for ease of notation we put $\hat{F}^{-1}(\alpha|M_t) \equiv \hat{y}_\alpha^{t+1}$ and $\hat{E}(Y_{t+1}|Y_{t+1} > F^{-1}(\alpha|M_t), M_t) \equiv S_\alpha^{t+1}$. We fix $m = 1000$, $n = 500$ and let $k = 100$ as suggested by the simulation study of McNeil and Frey (2000). Graph 3 provides the returns for one of our series (Dow 30 Industrial Stock price index against time) together with the VaR estimated by Nonparametric/L-M and GARCH-T/L-M methods. We note that GARCH-T/L-M seems to provide slightly smaller VaR estimates than Nonparametric/L-M estimator.

On backtesting the estimators of $\alpha-$VaR, we define a violation as the event $\{y_{t+1} > y_\alpha^{t+1}\}$. Under the null hypothesis that the return dynamics on y_t are correctly specified for a certain model/estimator, $I_t \equiv I\{y_{t+1} > y_\alpha^{t+1}\} \sim Bernoulli(1-\alpha)$ where $I(\cdot)$ is the indicator function. Consequently, $W = \sum_{t\in T} I_t \sim Binomial(N, 1-\alpha)$, where $N = Card(T)$ is the cardinality of set T. We perform a two sided test with the alternative hypothesis that the quantile is not correctly estimated with too many or too few violations. Under the null hypothesis the empirical version of the test statistic W, given by as $\hat{w} = \sum_{t\in T} I\{y_{t+1} > \hat{y}_\alpha^{t+1}\}$ is a drawing from a $Binomial(N, 1-\alpha)$. We report the violations numbers together with the p-values for different estimators in Table 4. The performance of different estimators are quite similar for the four historical series considered and the only rejection of the null hypothesis occurs for GARCH-T/ML method when estimating 99%-VaR for the Nasdaq series. Our nonparametric methods seem to outperform the others in estimating the Dow Jones' VaR and Microsoft's VaR for both values of α and in estimating 99%-VaR of Nasdaq.

In the case of $\alpha-$TailVaR, instead of using violation numbers to construct test statistics as in $\alpha-$VaR, we consider the normalized difference between y_{t+1} and S_α^{t+1},

$$r_{t+1} = \frac{y_{t+1} - E(Y_{t+1}|Y_{t+1} > F^{-1}(\alpha|M_t), M_t)}{\sigma_{t+1}} = \epsilon_{t+1} - E(\epsilon_{t+1}|\epsilon_{t+1} > q(\alpha)),$$

since $\alpha-$TailVaR gives the expected magnitude for returns given that violations have occured. If the return dynamics are correctly specified, given that $y_{t+1} > F^{-1}(\alpha|M_t)$, r_{t+1} is independent and identically distributed with mean zero. Hence, we use the estimated residuals $\{\hat{r}_{t+1} : t \in T \text{ and } y_{t+1} > \hat{y}_\alpha^{t+1}\}$, where

$$\hat{r}_{t+1} = \frac{y_{t+1} - S_\alpha^{t+1}}{\hat{\sigma}_{t+1}} = \frac{y_{t+1} - \hat{\mu}_{t+1}}{\hat{\sigma}_{t+1}} - \hat{E}(\epsilon_{t+1}|\epsilon_{t+1} > q(\alpha)).$$

Without making specific distribution assumptions on the residuals, we perform a one-sided bootstrap test as described in Efron and Tibshirani (1993, pp. 224-227) to test the null hypothesis that the mean of the residuals is zero against the alternative that the mean is greater than zero, since underestimating $\alpha-$TailVaR is likely to be the direction of interest. The p-values of the tests for different estimators are shown in Table 5. Given a 10% significance level for the test, our nonparametric/L-moments method for estimating $\alpha-$TailVaR is the only estimation procedure for which the null hypothesis is not rejected. In total, the null hypothesis for GARCH-T type estimators was rejected 5 out of 16 times, and for GARCH-N type estimators, the null hypothesis is rejected 9 out of 16 times. The empirical evidence seems to confirm the results from our Monte-Carlo study, supporting our conjecture that for the series considered in backtesting, nonlinearities may be more important to model than long memory in determing the stochastic structure of volatility.

6 Conclusion

In this paper we have proposed a novel way to estimate VaR and TailVaR, two measures of risk that have become extremely popular in the empirical, as well as theoretical finance literature. Our procedure combines the methodology originally proposed by McNeil and Frey (2000) with nonparametric models of volatility dynamics and L-moment estimation. A Monte Carlo study that is based on a DGP that incorporates empirical regularities of returns on financial time series reveals that our estimation method outperforms the methodology put forth by McNeil and Frey. To the best of our knowledge, this is the first evidence on the finite sample performance of VaR and TailVaR estimators for conditional densities. It is important at this point to highlight the fact that asymptotic results for these types of estimators are currently unavailable. In addition, results from our simulations seem to indicate that nonlinearities in volatility dynamics may be very important in accurately estimating measures of risk. In fact, our simulations indicate that accounting for nonlinearities may be more important than richer modeling of dependency. Our proposed estimation procedure was also evaluated *via* backtesting, and the results, although specific to the series analyzed seem to confirm the more general conclusions of the Monte Carlo.

Appendix 1

Proposition 1 : Let $g(y;v) = c\left(1+\frac{1}{v-2}y^2\right)^m$ where $c, m \in \Re$, $v > 2$ a positive integer and $-\infty < y < \infty$. Let $\kappa_p = \int_0^\infty y^p g(y;v)dy$ for $p = 1,2,3,4$. Then,

$\kappa_1 = \frac{c(v-2)}{2}\int_0^1 u^{-m-2}du,$
$\kappa_2 = \frac{c(v-2)^{3/2}}{2}\left(B[1/2,-m-3/2] - B[1/2,-m-1/2]\right),$
$\kappa_3 = \frac{c(v-2)^2}{2}\left(\int_0^1 u^{-m-3}du - \int_0^1 u^{-m-2}du\right),$
$\kappa_4 = \frac{c(v-2)^{5/2}}{2}\left(B[1/2,-m-5/2] - 2B[1/2,-m-3/2] + B[1/2,-m-1/2]\right),$

where $B[\alpha,\beta]$ is the beta function.

Proof: Let $\theta = (v-2)^{-1/2}y$, then $\kappa_1 = c(v-2)\int_0^\infty \theta(1+\theta^2)^m d\theta$. Now, put $\theta = tan(w)$ and using the fact $sin(w)^2 = 1 - cos(w)^2$ and $1 + tan^2(w) = cos(w)^{-2}$, we have

$$\kappa_1 = -c(v-2)\int_0^{\pi/2} cos(w)^{-2m-3}dcos(w) = \frac{c(v-2)}{2}\int_0^1 u^{-m-2}du.$$

For κ_2 we have, $\kappa_2 = c(v-2)^{3/2}\int_0^\infty \theta^2(1+\theta^2)^m d\theta$. Using the same transformations above, we have

$$\kappa_2 = c(v-2)^{3/2}\left(\int_0^{\pi/2} cos(w)^{-2m-4}dw - \int_0^{\pi/2} cos(w)^{-2m-2}dw\right).$$

It is easy to show that for $h \in \Re$ and $B[\alpha,\beta]$ the beta function,

$$\begin{aligned}\int_0^{\pi/2} cos(w)^{-2m-h}dw &= \frac{1}{2}\int_0^1 u^{1/2-1}(1-u)^{-m-h/2-1/2}du \\ &= \frac{1}{2}B[1/2,-m-h/2+1/2],\end{aligned}$$

which gives the desired result. For κ_3 we have, $\kappa_3 = c(v-2)^2\int_0^\infty \theta^3(1+\theta^2)^m d\theta$ and we obtain

$$\kappa_3 = c(v-2)^2\left(\int_0^{\pi/2} cos(w)^{-2m-3}dcos(w) - \int_0^{\pi/2} cos(w)^{-2m-5}dcos(w)\right)$$

$$= \frac{c(v-2)^2}{2}\left(\int_0^1 u^{-m-3}du - \int_0^1 u^{-m-2}du\right)$$

Finally, for κ_4 we have $\kappa_4 = c(v-2)^{5/2}\int_0^\infty \theta^4(1+\theta^2)^m d\theta$ and it is straightforward to show that,

$$\kappa_4 = c(v-2)^{5/2}\int_0^{\pi/2}\left(cos(w)^{-2m-6} - 2cos(w)^{-2m-4} + cos(w)^{-2m-2}\right)dw$$

Using the previous results we obtain the desired expression.

Proof (Lemma 1): From Hansen (1994), $\int_{-\infty}^{\infty} xf(x;v,\lambda)dx = 0$ and

$$\int_{-\infty}^{\infty} x^2 f(x;v,\lambda)dx = 1,$$

therefore $\alpha_3 = \int_{-\infty}^{\infty} x^3 f(x;v,\lambda)dx$ and $\alpha_4 = \int_{-\infty}^{\infty} x^4 f(x;v,\lambda)dx$. First consider α_3. Note that

$$\begin{aligned}\alpha_3 &= \int_{-\infty}^{-a/b} x^3 bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1-\lambda}\right)^2\right)^{-(v+1)/2} dx \\ &+ \int_{-a/b}^{\infty} x^3 bc\left(1+\frac{1}{v-2}\left(\frac{bx+a}{1+\lambda}\right)^2\right)^{-(v+1)/2} dx.\end{aligned}$$

Let $y = \frac{bx+a}{1-\lambda}$ on the first integral and $y = \frac{bx+a}{1+\lambda}$ on the second integral. Then,

$$\begin{aligned}\alpha_3 &= \int_{-\infty}^{0}\left(\frac{1-\lambda}{b}y - a/b\right)^3 c\left(1+\frac{1}{v-2}y^2\right)^{-\frac{v+1}{2}}(1-\lambda)dy \\ &+ \int_{0}^{\infty}\left(\frac{1+\lambda}{b}y - a/b\right)^3 c\left(1+\frac{1}{v-2}y^2\right)^{-\frac{v+1}{2}}(1+\lambda)dy.\end{aligned}$$

Simple manipulations yield, $\alpha_3 = \frac{8\kappa_3}{b^3}(\lambda^3+\lambda) - \frac{6\kappa_2 a}{b^3}(\lambda^3+3\lambda) + \frac{12a^2\kappa_1}{b^3}\lambda - \frac{a^3+3a(1-\lambda)^3}{b^3}$, where κ_i for $i = 1,2,3,4$ are as defined in Proposition 1. Using the same transformations for α_4, we have

$$\begin{aligned}\alpha_4 &= \frac{2\lambda^5+20\lambda^3+10\lambda}{b^4}(\kappa_4) - \frac{32a\kappa_3}{b^4}(\lambda^3+\lambda) + \frac{12a^2\kappa_2}{b^4}(\lambda^3+3\lambda) - \frac{16a^3\lambda\kappa_1}{b^4} \\ &+ \frac{1}{b^4}\left(a^4 + \frac{3(1-\lambda)^5(v-2)}{v-4} + 6a^2(1-\lambda)^3\right)\end{aligned}$$

Proof (Lemma 2):

$$E(X|X>z)=\int_z^\infty x f_{X>z}(x;v,\lambda)dx=(1-F(z))^{-1}\int_z^\infty x f(x;v,\lambda)dx.$$

Let $I=\int_z^\infty x f(x;v,\lambda)dx$ and put $y=\frac{bx+a}{1+\lambda}$, then

$$I=\frac{(1+\lambda)^2}{b}\int_\alpha^\infty x g(x;v)dx-\frac{(1+\lambda)a}{b}\int_\alpha^\infty g(x;v)dx \tag{6.1}$$

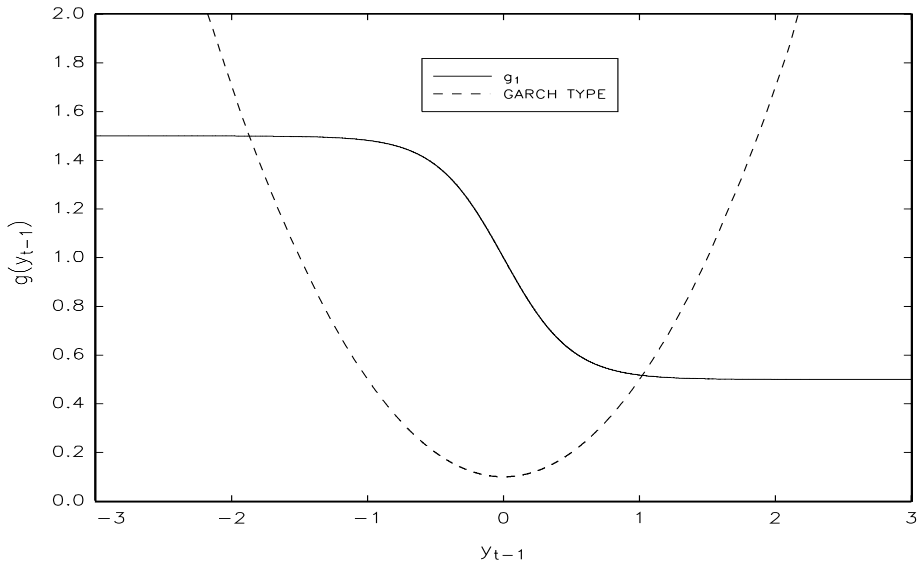
where $\alpha=\frac{bz+a}{1+\lambda}$ and $g(x;v)=c\left(1+\frac{1}{v-2}x^2\right)^{-(v+1)/2}$. From Lemma 1,

$$\int_\alpha^\infty x g(x;v)dx=\frac{c(v-2)}{2}\int_0^{cos^2(arctan(\gamma))}u^{(v+1)/2-2}du \tag{6.2}$$

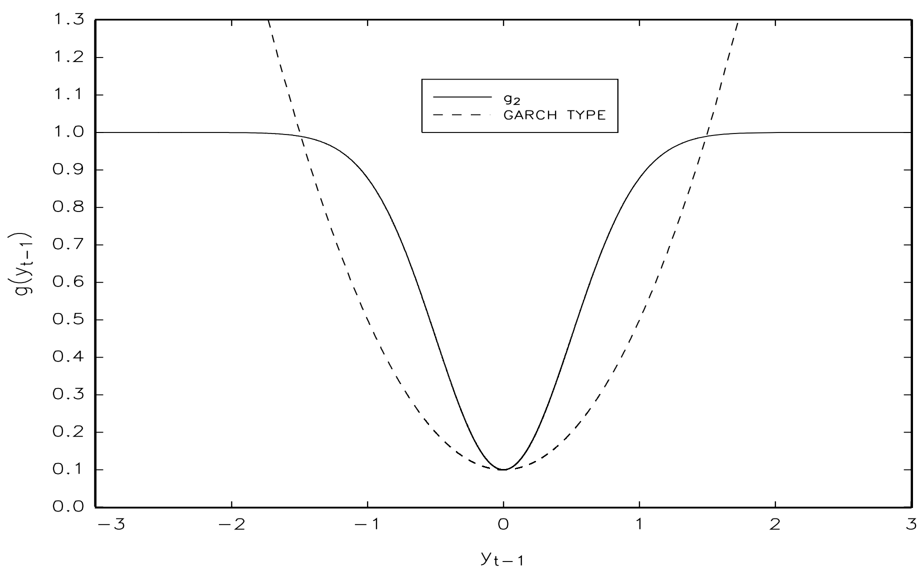
where $\gamma=(v-2)^{-0.5}\alpha$. Consequently, $\int x g(x;v)dx=\frac{c(v-2)}{v-1}\beta^{\frac{v-1}{2}}$. For the second integral note that from Lemma 1, it is easy to show that $\int_\alpha^\infty g(x;v)dx=1-\int_{-\infty}^\alpha g(x;v)dx=1-F_s\left(\frac{bz+a}{1+\lambda}\sqrt{\frac{v}{v-2}}\right)$, which combined with (6.2) and substituting back in (6.1) gives the desired expression.

Appendix 2 - Tables and Graphs

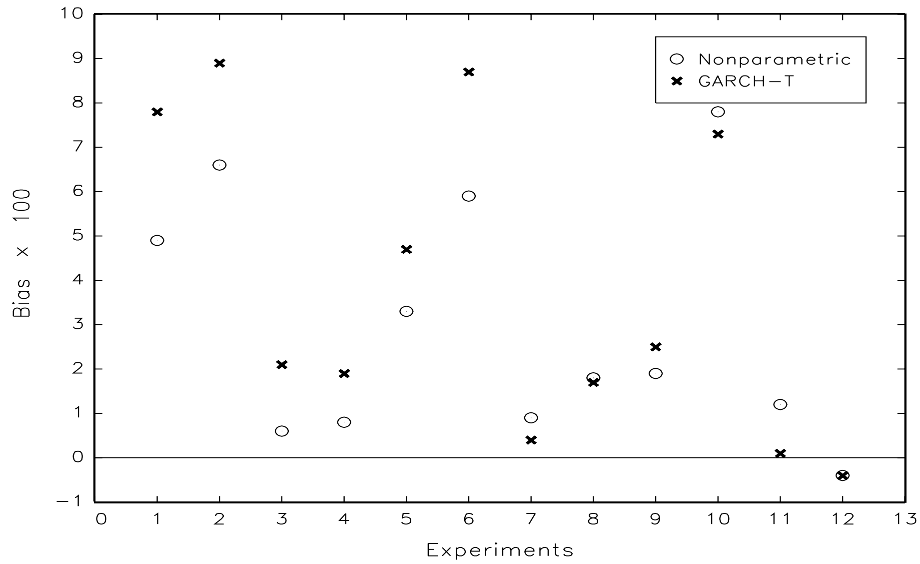
Graph 1a conditional volatility based on $g_1(x)$ and GARCH Model



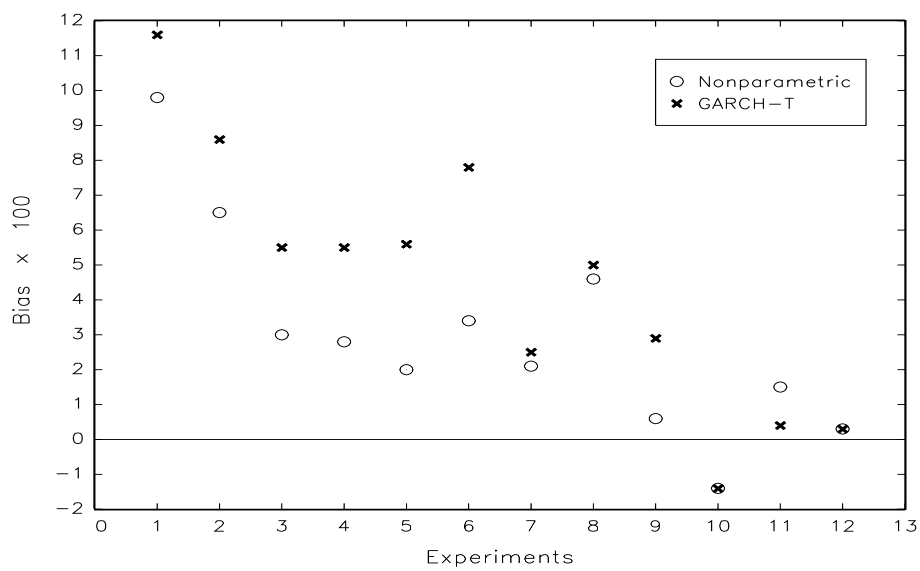
Graph 1b conditional volatility based on $g_2(x)$ and GARCH Model



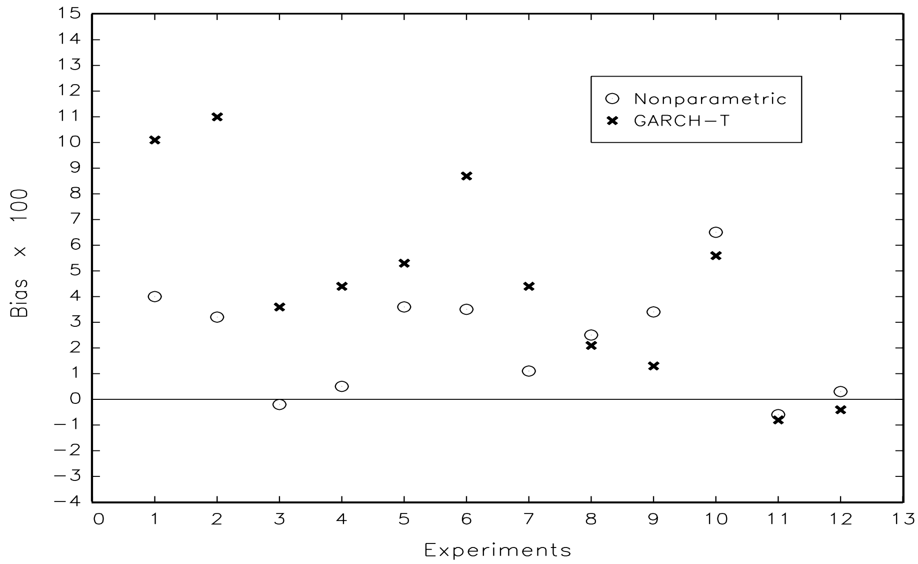
GRAPH 2A BIAS $\times 100$ ON VaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_1(x)$ FOR GARCH-T, NONPARAMETRIC



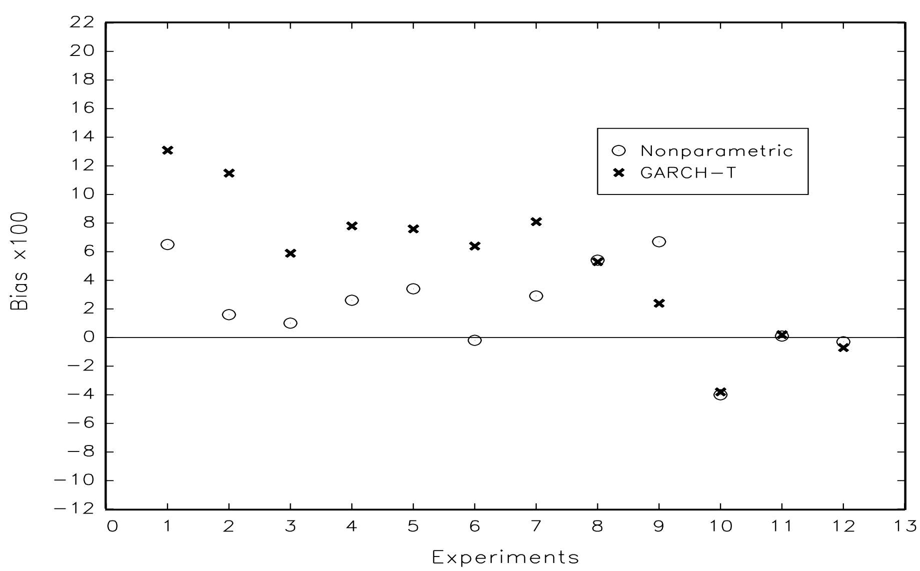
GRAPH 2B BIAS $\times 100$ ON TailVaR USING L-MOMENTS WITH $n = 1000$, $\lambda = -0.25$, VOLATILITY BASED ON $g_1(x)$ FOR GARCH-T, NONPARAMETRIC



Graph 2c Bias $\times 100$ on VaR using L-Moments with $n = 1000$, $\lambda = -0.25$, Volatility based on $g_2(x)$ for GARCH-t, Nonparametric



Graph 2d Bias $\times 100$ on TailVaR using L-Moments with $n = 1000$, $\lambda = -0.25$, Volatility based on $g_2(x)$ for GARCH-t, Nonparametric



Graph 3 Returns for Dow 30 Industrial Stock price index, VaR is estimated by Nonparametric/L-M and GARCH-T/L-M methods

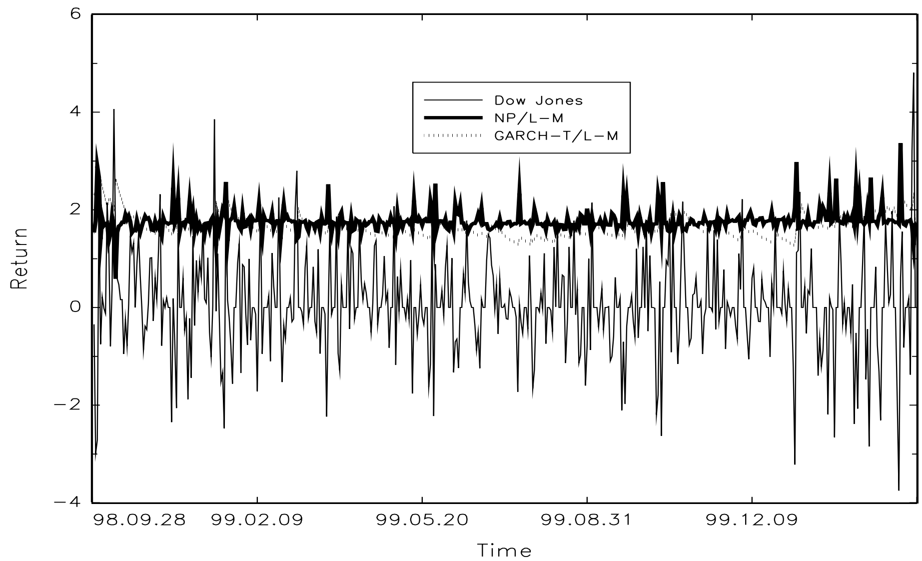


TABLE 1 NUMBERING OF EXPERIMENTS
$\lambda = 0,\ -0.25,\ -0.5$; $n = 500,\ 1000$;VOLATILITY BASED ON $g_1(x),\ g_2(x)$

Exp	γ	α	k
1	0.3	0.99	60
2	0.3	0.99	100
3	0.3	0.95	60
4	0.3	0.95	100
5	0.6	0.99	60
6	0.6	0.99	100
7	0.6	0.95	60
8	0.6	0.95	100
9	0.9	0.99	60
10	0.9	0.99	100
11	0.9	0.95	60
12	0.9	0.95	100

Table 2A Relative MSE for $n = 1000,\ \lambda = 0,$
Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.322	1.429	2.360	2.348	2.790	2.888	2.888	3.058	2.838	2.957	2.882	3.076
2	1.290	1.285	1.964	1.783	2.538	2.531	2.588	2.552	2.552	2.572	2.493	2.484
3	1.063	1.088	1.285	1.294	2.781	2.783	2.502	2.522	2.821	2.818	2.500	2.538
4	1.056	1.038	1.305	1.218	2.586	2.599	2.531	2.543	2.752	2.753	2.688	2.706
5	1.117	1.178	1.472	1.572	1.528	1.613	1.726	1.926	1.511	1.602	1.685	1.874
6	1.179	1.180	1.653	1.590	1.457	1.513	1.675	1.757	1.519	1.555	1.685	1.749
7	1.057	1.060	1.088	1.099	1.409	1.415	1.412	1.431	1.549	1.559	1.548	1.591
8	1.078	1.062	1.189	1.141	1.599	1.596	1.611	1.624	1.550	1.545	1.558	1.568
9	1.159	1.262	1.500	1.788	1.076	1.210	1.591	1.928	1.193	1.329	1.691	1.973
10	1.145	1.192	1.547	1.609	0.983	1.051	1.504	1.598	1.249	1.273	1.707	1.754
11	1.073	1.114	1.111	1.265	0.951	0.982	0.995	1.099	1.024	1.049	1.078	1.192
12	1.026	1.039	1.032	1.030	1.034	1.043	1.008	1.004	1.106	1.116	1.110	1.107

Table 2B Relative MSE for $n = 1000$, $\lambda = -0.25$, Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.284	1.384	2.271	2.177	2.481	2.597	2.409	2.534	2.694	2.802	2.526	2.668
2	1.225	1.247	1.656	1.541	2.729	2.720	2.510	2.462	2.927	2.933	2.605	2.583
3	1.081	1.091	1.257	1.379	2.838	2.844	2.771	2.875	3.211	3.214	3.112	3.188
4	1.032	1.035	1.187	1.173	2.861	2.863	2.943	2.966	3.146	3.139	3.186	3.179
5	1.125	1.214	1.378	1.505	1.511	1.576	1.724	1.861	2.154	2.197	2.282	2.355
6	1.140	1.147	1.517	1.499	1.776	1.808	1.899	1.912	1.885	1.891	2.064	2.056
7	1.021	1.031	1.080	1.101	1.396	1.401	1.375	1.377	1.791	1.807	1.752	1.781
8	1.081	1.094	1.236	1.288	1.648	1.679	1.660	1.751	1.783	1.813	1.780	1.871
9	1.096	1.266	1.381	1.612	1.148	1.245	1.408	1.567	1.253	1.368	1.502	1.678
10	1.184	1.230	1.573	1.603	1.043	1.087	1.447	1.528	1.304	1.337	1.688	1.743
11	1.097	1.103	1.058	1.150	1.088	1.098	1.002	1.061	1.178	1.187	1.089	1.180
12	1.049	1.060	1.101	1.122	1.063	1.073	1.052	1.098	1.084	1.096	1.055	1.090

Table 2C Relative MSE for $n = 1000,\ \lambda = -0.5$,
Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.177	1.209	1.768	1.706	2.775	2.848	2.766	2.881	3.470	3.516	3.284	3.387
2	1.209	1.226	1.395	1.377	1.393	1.401	1.413	1.418	1.427	1.443	1.435	1.447
3	1.003	1.009	1.199	1.206	2.501	2.510	2.322	2.360	3.239	3.247	2.971	3.020
4	1.017	1.042	1.169	1.202	2.797	2.791	2.734	2.754	3.714	3.681	3.500	3.493
5	1.062	1.142	1.296	1.475	1.438	1.553	1.453	1.625	1.700	1.792	1.647	1.796
6	1.081	1.126	1.224	1.223	1.392	1.411	1.350	1.367	1.624	1.639	1.505	1.523
7	1.006	1.012	1.046	1.050	1.355	1.355	1.297	1.322	1.610	1.617	1.521	1.549
8	1.011	1.021	1.099	1.124	1.136	1.127	1.138	1.137	1.319	1.312	1.308	1.313
9	1.087	1.264	1.830	1.754	1.010	1.150	1.349	1.579	1.728	1.798	1.910	2.036
10	1.078	1.128	1.303	1.345	0.946	0.969	1.169	1.177	1.088	1.094	1.270	1.256
11	1.006	1.012	1.009	1.044	0.788	0.795	0.661	0.731	0.893	0.897	0.731	0.786
12	0.962	0.969	0.974	0.995	0.935	0.940	0.866	0.881	1.452	1.445	1.313	1.337

TABLE 2D RELATIVE MSE FOR $n = 1000,\ \lambda = 0$,
VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.077	1.107	1.370	1.354	3.492	3.570	3.453	3.440	3.805	3.912	3.653	3.645
2	1.073	1.125	1.326	1.344	6.857	6.852	6.368	6.337	8.578	8.603	7.939	7.951
3	0.987	0.996	1.051	1.077	4.169	4.191	4.428	4.287	5.030	5.056	5.274	5.134
4	1.017	1.025	1.065	1.066	3.596	3.645	3.449	3.441	4.295	4.345	4.011	3.996
5	1.039	1.058	1.255	1.353	1.662	1.667	1.886	1.897	1.885	1.918	1.971	2.081
6	1.062	1.073	1.358	1.388	1.740	1.753	1.977	2.031	2.151	2.201	2.270	2.425
7	1.013	1.021	1.033	1.066	3.221	3.242	3.116	3.148	2.376	2.391	2.309	2.353
8	1.000	1.000	1.021	1.018	1.192	1.199	1.237	1.245	1.368	1.376	1.413	1.428
9	1.104	1.256	1.404	1.695	0.869	0.997	1.354	1.604	1.068	1.188	1.481	1.752
10	1.102	1.162	1.381	1.457	0.980	0.997	1.384	1.390	1.065	1.083	1.434	1.423
11	1.070	1.067	1.040	1.101	0.819	0.836	0.867	1.005	1.043	1.051	1.052	1.164
12	1.044	1.041	1.055	1.069	0.738	0.754	0.772	0.804	0.920	0.922	0.943	0.947

Table 2E Relative MSE for $n = 1000,\ \lambda = -0.25$,
Volatility based on $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	1.092	1.116	1.371	1.390	6.007	6.054	5.471	5.392	7.197	7.289	6.495	6.318
2	1.054	1.061	1.245	1.257	3.945	3.958	3.576	3.635	4.650	4.688	4.096	4.195
3	1.019	1.018	1.035	1.052	3.702	3.736	3.724	3.750	3.819	3.856	3.771	3.789
4	1.010	1.013	1.077	1.067	3.026	3.040	3.227	3.226	3.753	3.792	3.855	3.856
5	1.060	1.111	1.308	1.378	1.273	1.336	1.505	1.567	1.606	1.694	1.772	1.866
6	1.070	1.083	1.270	1.229	2.071	2.067	2.053	2.043	3.329	3.336	3.036	3.039
7	1.032	1.034	1.084	1.115	1.844	1.851	1.967	1.991	3.219	3.230	3.135	3.159
8	1.039	1.035	1.120	1.112	1.482	1.506	1.515	1.494	1.807	1.822	1.811	1.816
9	1.132	1.264	3.247	2.039	0.929	1.141	1.476	1.852	0.995	1.187	1.500	1.834
10	1.089	1.119	1.390	1.405	0.870	0.903	1.253	1.256	0.936	1.002	1.255	1.309
11	1.014	1.025	1.019	1.115	0.850	0.858	0.833	0.912	0.989	0.999	0.939	1.001
12	1.010	1.016	1.073	1.079	0.803	0.800	0.856	0.888	1.156	1.159	1.140	1.143

Table 2F Relative MSE for $n = 1000,\ \lambda = -0.5$,
Volatility based on $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE	L-M	MLE
1	0.991	1.022	1.120	1.164	4.309	4.347	3.586	3.614	5.098	5.135	4.180	4.233
2	1.086	1.089	1.415	1.200	3.326	3.341	2.877	2.878	3.884	3.894	3.362	3.335
3	1.005	1.007	1.055	1.075	7.276	7.304	7.219	7.202	11.126	11.087	10.536	10.201
4	0.989	0.988	1.068	1.030	3.886	3.895	3.753	3.772	4.816	4.836	4.614	4.627
5	1.044	1.073	1.178	1.212	1.358	1.395	1.364	1.403	2.099	2.117	1.980	1.988
6	1.086	1.085	1.198	1.180	1.419	1.425	1.327	1.336	1.900	1.916	1.668	1.692
7	1.009	1.020	1.030	1.053	1.478	1.478	1.411	1.427	2.047	2.058	1.899	1.918
8	1.008	1.016	1.074	1.078	1.413	1.425	1.323	1.332	2.085	2.110	1.921	1.924
9	1.070	1.129	1.441	1.512	0.748	0.848	1.087	1.273	1.039	1.085	1.302	1.433
10	0.928	0.931	1.010	0.982	0.663	0.675	0.861	0.846	0.780	0.785	0.942	0.908
11	0.987	0.996	1.278	1.311	0.782	0.789	0.745	0.797	0.900	0.908	0.830	0.890
12	1.019	1.016	1.093	1.090	0.787	0.806	0.770	0.835	1.718	1.711	0.514	1.579

Table 3A MSE for $\lambda = 0$,
Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.219	0.134	0.805	0.572	0.350	0.283	0.965	0.700	0.354	0.288	0.964	0.699
2	0.197	0.139	0.696	0.510	0.306	0.273	0.767	0.672	0.333	0.274	0.844	0.647
3	0.055	0.034	0.149	0.098	0.107	0.088	0.242	0.190	0.113	0.089	0.254	0.190
4	0.049	0.035	0.143	0.100	0.098	0.086	0.221	0.194	0.102	0.091	0.231	0.205
5	0.349	0.220	1.206	0.678	0.405	0.301	1.366	0.795	0.444	0.297	1.418	0.776
6	0.343	0.238	1.145	0.775	0.367	0.294	1.067	0.785	0.401	0.307	1.121	0.790
7	0.086	0.068	0.217	0.156	0.115	0.091	0.286	0.203	0.122	0.100	0.291	0.223
8	0.094	0.065	0.275	0.167	0.115	0.097	0.292	0.226	0.123	0.094	0.313	0.219
9	1.330	0.619	5.870	2.137	1.012	0.575	5.375	2.267	1.093	0.637	5.525	2.409
10	1.145	0.614	3.589	2.080	0.987	0.527	3.360	2.022	1.016	0.670	3.426	2.296
11	0.303	0.168	0.797	0.441	0.251	0.149	0.637	0.395	0.252	0.161	0.641	0.428
12	0.266	0.165	0.792	0.424	0.230	0.167	0.595	0.414	0.244	0.178	0.658	0.456

Table 3B MSE for $\lambda = -0.25$,
Volatility based on $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.120	0.088	0.442	0.334	0.200	0.170	0.486	0.354	0.217	0.185	0.498	0.371
2	0.133	0.078	0.370	0.239	0.206	0.175	0.467	0.363	0.230	0.187	0.488	0.376
3	0.041	0.022	0.111	0.052	0.082	0.058	0.166	0.115	0.107	0.066	0.217	0.129
4	0.042	0.022	0.107	0.051	0.072	0.061	0.149	0.126	0.162	0.067	0.297	0.137
5	0.207	0.144	0.796	0.366	0.240	0.194	0.611	0.458	0.337	0.277	0.741	0.607
6	0.198	0.131	0.543	0.369	0.235	0.204	0.592	0.462	0.258	0.216	0.604	0.502
7	0.074	0.046	0.156	0.104	0.086	0.064	0.179	0.132	0.093	0.081	0.194	0.168
8	0.061	0.043	0.142	0.100	0.098	0.066	0.198	0.134	0.117	0.072	0.232	0.144
9	0.610	0.341	2.143	1.067	0.526	0.357	1.888	1.087	0.585	0.390	1.969	1.160
10	0.696	0.357	2.044	1.097	0.606	0.314	1.756	1.009	2.193	0.393	3.820	1.178
11	0.168	0.110	0.450	0.231	0.150	0.110	0.354	0.219	0.166	0.119	0.380	0.238
12	0.187	0.129	0.436	0.286	0.179	0.130	0.406	0.273	0.220	0.133	0.496	0.274

TABLE 3C MSE FOR $\lambda = -0.5$,
VOLATILITY BASED ON $g_1(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.072	0.039	0.204	0.105	0.145	0.091	0.247	0.164	0.475	0.114	0.631	0.194
2	0.086	0.096	0.219	0.183	0.123	0.110	0.201	0.186	0.143	0.113	0.218	0.188
3	0.026	0.020	0.061	0.043	0.046	0.051	0.079	0.084	0.059	0.066	0.102	0.107
4	0.029	0.016	0.076	0.031	0.047	0.043	0.079	0.072	0.057	0.057	0.091	0.093
5	0.104	0.072	0.322	0.172	0.123	0.098	0.248	0.192	0.161	0.115	0.300	0.218
6	0.116	0.083	0.275	0.167	0.126	0.106	0.237	0.184	0.182	0.124	0.297	0.206
7	0.042	0.034	0.079	0.063	0.058	0.046	0.095	0.078	0.091	0.054	0.149	0.091
8	0.049	0.041	0.108	0.075	0.074	0.046	0.120	0.078	0.106	0.053	0.179	0.089
9	0.297	0.180	0.865	0.581	0.242	0.168	0.661	0.428	0.337	0.287	0.808	0.606
10	0.442	0.205	1.067	0.496	0.272	0.180	0.685	0.445	0.360	0.207	0.788	0.483
11	0.103	0.095	0.230	0.202	0.094	0.075	0.179	0.132	0.137	0.085	0.245	0.146
12	0.143	0.073	0.257	0.133	0.115	0.071	0.201	0.119	0.124	0.110	0.213	0.180

TABLE 3D MSE FOR $\lambda = 0$,
VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.182	0.126	0.461	0.319	0.632	0.410	1.616	0.803	0.410	0.446	0.942	0.849
2	0.196	0.119	0.514	0.293	0.367	0.763	0.741	1.408	0.579	0.954	1.326	1.755
3	0.052	0.033	0.117	0.074	0.160	0.140	0.359	0.313	0.186	0.169	0.394	0.372
4	0.050	0.035	0.121	0.081	0.184	0.123	0.385	0.261	0.217	0.147	0.429	0.304
5	0.302	0.259	0.767	0.665	0.372	0.414	0.995	1.000	0.646	0.470	1.673	1.045
6	0.318	0.253	0.796	0.660	0.506	0.415	1.196	0.961	0.530	0.512	1.193	1.103
7	0.103	0.075	0.239	0.167	0.120	0.238	0.284	0.504	0.138	0.176	0.313	0.374
8	0.104	0.077	0.242	0.166	0.202	0.091	0.450	0.201	0.219	0.105	0.467	0.230
9	1.256	0.551	4.547	1.802	0.708	0.434	2.882	1.737	0.894	0.533	3.148	1.901
10	1.004	0.557	3.801	1.840	0.765	0.495	2.702	1.845	0.882	0.538	2.954	1.911
11	0.269	0.166	0.798	0.368	0.178	0.127	0.532	0.307	0.220	0.162	0.615	0.372
12	0.247	0.143	0.609	0.368	0.164	0.101	0.477	0.270	0.238	0.126	0.630	0.329

Table 3E MSE for $\lambda = -0.25$,
Volatility based on $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.099	0.070	0.396	0.165	0.475	0.386	0.919	0.659	0.479	0.463	0.903	0.783
2	0.126	0.075	0.386	0.164	0.308	0.281	0.516	0.471	0.348	0.332	0.557	0.540
3	0.036	0.027	0.075	0.053	0.149	0.098	0.283	0.191	0.179	0.102	0.326	0.193
4	0.032	0.029	0.069	0.060	0.150	0.087	0.276	0.180	0.146	0.107	0.267	0.215
5	0.222	0.152	0.568	0.356	0.249	0.183	0.601	0.410	0.412	0.231	0.786	0.482
6	0.228	0.149	0.491	0.334	0.325	0.288	0.635	0.540	0.612	0.463	1.154	0.798
7	0.070	0.057	0.146	0.114	0.097	0.102	0.195	0.208	0.123	0.179	0.238	0.331
8	0.077	0.056	0.163	0.117	0.093	0.080	0.186	0.158	0.134	0.097	0.251	0.189
9	0.509	0.294	1.577	1.969	0.443	0.241	1.544	0.895	0.521	0.258	1.634	0.910
10	0.551	0.290	1.750	0.935	0.458	0.232	1.508	0.843	0.574	0.249	1.631	0.844
11	0.158	0.095	0.365	0.207	0.144	0.080	0.316	0.169	0.156	0.093	0.333	0.191
12	0.155	0.090	0.362	0.201	0.116	0.071	0.290	0.160	0.767	0.103	1.479	0.214

TABLE 3F MSE FOR $\lambda = -0.5$,
VOLATILITY BASED ON $g_2(x)$

Exp	Nonparametric				GARCH-T				GARCH-N			
	VaR		TailVaR		VaR		TailVaR		VaR		TailVaR	
	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000	n= 500	n= 1000
1	0.062	0.044	0.131	0.089	0.137	0.192	0.214	0.284	0.160	0.227	0.235	0.330
2	0.068	0.041	0.135	0.095	0.186	0.124	0.251	0.194	0.234	0.145	0.308	0.227
3	0.025	0.019	0.046	0.033	0.077	0.137	0.125	0.227	0.079	0.210	0.126	0.331
4	0.026	0.017	0.046	0.033	0.104	0.068	0.161	0.114	0.121	0.084	0.187	0.141
5	0.133	0.088	0.281	0.164	0.160	0.115	0.284	0.190	0.227	0.177	0.365	0.275
6	0.121	0.099	0.255	0.194	0.161	0.129	0.266	0.215	0.206	0.173	0.310	0.270
7	0.045	0.039	0.092	0.068	0.056	0.057	0.093	0.094	0.108	0.080	0.169	0.126
8	0.049	0.044	0.093	0.081	0.063	0.062	0.101	0.100	0.095	0.091	0.146	0.145
9	0.236	0.179	1.734	0.458	0.208	0.125	0.569	0.346	0.370	0.174	0.791	0.414
10	0.288	0.167	1.262	0.394	0.193	0.119	0.621	0.336	0.219	0.140	0.630	0.368
11	0.086	0.062	0.168	0.150	0.082	0.049	0.149	0.087	0.197	0.057	0.327	0.097
12	0.095	0.062	0.194	0.119	0.066	0.048	0.128	0.084	0.085	0.104	0.161	0.165

Table 4 Backtest results for α-VaR
numbers of violations and p-value (in brackets)
test length $m-n=500$, expected violations $=(m-n)(1-\alpha)$

$\alpha=0.95$	Nonparametric		GARCH-T		GARCH-N	
	L-M	MLE	L-M	MLE	L-M	MLE
Expect	25		25		25	
Dow Jones	29 (0.412)	30 (0.305)	32 (0.151)	33 (0.101)	31 (0.218)	31 (0.218)
Microsoft	23 (0.682)	24 (0.837)	22 (0.538)	24 (0.837)	18 (0.151)	19 (0.218)
Nasdaq	21 (0.412)	19 (0.218)	23 (0.682)	22 (0.538)	21 (0.412)	20 (0.305)
S&P500	21 (0.412)	21 (0.412)	21 (0.412)	20 (0.305)	23 (0.682)	23 (0.682)

$\alpha=0.99$	Nonparametric		GARCH-T		GARCH-N	
	L-M	MLE	L-M	MLE	L-M	MLE
Expect	5		5		5	
Dow Jones	6 (0.653)	6 (0.653)	7 (0.369)	7 (0.369)	6 (0.653)	6 (0.653)
Microsoft	5 (1.000)	5 (1.000)	7 (0.369)	6 (0.653)	7 (0.369)	7 (0.369)
Nasdaq	5 (1.000)	5 (1.000)	8 (0.178)	9 (0.072)	8 (0.178)	8 (0.178)
S&P500	4 (0.653)	3 (0.369)	6 (0.653)	6 (0.653)	6 (0.653)	6 (0.653)

Table 5 Backtest results for α-TailVaR
p-values of one-sided Bootstrap test

$\alpha = 0.95$	Nonparametric		GARCH-T		GARCH-N	
	L-M	MLE	L-M	MLE	L-M	MLE
Dow Jones	0.464	0.522	0.490	0.533	0.348	0.327
Microsoft	0.994	0.994	0.053	0.129	0.023	0.059
Nasdaq	0.248	0.082	0.113	0.033	0.030	0.010
S&P500	0.550	0.625	0.014	0.002	0.060	0.028
$\alpha = 0.99$	Nonparametric		GARCH-T		GARCH-N	
	L-M	MLE	L-M	MLE	L-M	MLE
Dow Jones	0.108	0.123	0.333	0.309	0.100	0.121
Microsoft	0.937	0.942	0.137	0.136	0.131	0.230
Nasdaq	0.108	0.072	0.117	0.052	0.045	0.014
S&P500	0.312	0.129	0.140	0.103	0.120	0.070

References

Aït-Sahalia, Y. and M.W. Brandt (2001): "Variable selection for portfolio choice," *Journal of Finance*, 56, 1297-1355.

Andreou, E., N. Pittis and A. Spanos (2001): "On modelling speculative prices: the empirical literature," *Journal of Economic Surveys*, 15, 187-220.

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

Basle Committee (1996): *Overview of the ammendment of the capital accord to incorporate market risk*. Basle Committee on Banking Supervision.

Bollerslev, T. (1986): "Generalized autoregressive conditional heterocedasticity," *Journal of Econometrics*, 31, 307-327.

Carroll, R.J., W. Härdle and E. Mammen (2002): "Estimation in an additive model when the components are linked parametrically," *Econometric Theory*, 18, 886-912.

Chen, Yi-Ting (2001): "Testing conditional symmetry with an application to stock returns," Working Paper, Academia Sinica.

Diebold, F.X., T. Schuermann, J.D. Stroughair (1998): "Pitfalls and opportunities in the use of extreme value theory in risk management," in A.-P. Refenes, J.D. Moody and A.N. Burgess eds., *Advances in Computational Finance*, Amsterdam: Kluwer Academic, 2-12.

Diebolt, J. and D. Guégan (1993): "Tail behaviour of the stationary density of general nonlinear autoregressive processes of order 1," *Journal of Applied Probability*, 30, 315-329.

Ding, Z., C.W.J. 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.

Efron B. and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*. New York: Chapman and Hall/CRC.

Embrechts, P., S. Resnick, and G. Samorodnitsky (1999): "Extreme value theory as a risk management tool," *North American Actuarial Journal*, 3, 30-41.

– –, C. Klüppelberg, T. Mikosh (1997): *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Fan, J. (1992): "Design adaptive nonparametric regression," *Journal of the American Statistical Association*, 87, 998-1004.

– – and Q. Yao (1998): "Efficient estimation of conditional variance functions in stochastic regression," *Biometrika*, 85, 645-660.

Gouriéroux, C. (1997): *ARCH Models and Financial Applications*. Berlin: Springer.

Hafner, C. (1998): *Nonlinear Time Series Analysis with Applications to Foreign Exchange Volatility*. Heidelberg: Physica-Verlag.

Hannan, E. J. (1987): "The Cochrane-Orcutt papers," in *Specification Analysis in the Linear Model*, London: Routledge and Kegan Paul, 9-18.

Hansen, B. (1994): "Autoregressive conditional density estimation," *International Economic Review*, 35, 705-729.

Härdle, W. and A. Tsybakov (1997): "Local Polynomial estimators of the volatility function in nonparametric autoregression," *Journal of Econometrics*, 81, 223-242.

Hosking, J. R. M. (1987): "Parameter and quantile estimation for the generalized pareto distribution," *Technometrics*, 29, 339-349.

– – (1989): "Some theoretical results concerning L-moments," IBM Research Division, Technical Report.

– – (1990): "L-moments: analysis and estimation of distributions using linear combinations of order statistics," *Journal of the Royal Statistical Society B*, 52, 105-124.

– – and J. R. Wallis (1997): *Regional Frequency Analysis*. Cambridge: Cambridge University Press.

– –, – – and E. F. Wood (1985): "Estimation of the generalized extreme value distribution by the method of probability weighted moments," *Technometrics,* 27, 251-261.

Linton, O. and J. P. Nielsen (1995): "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika*, 82, 1, 93-100.

Longin, F. (2000): "From value-at-risk to stress testing, the Extreme Value Approach," *Journal of Banking and Finance*, 24, 1097-1130.

Mandy, D. and C. Martins-Filho (2001): "Optimal IV estimation of systems with stochastic regressors and VAR disturbances with applications to dynamic systems," *Econometric Reviews*, 20, 485-505.

Masry, E. and D. Tjøstheim (1995): "Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality," *Econometric Theory*, 11, 258-289.

McNeil, A.J. and R. Frey (2000): "Estimation of tail-related risk measures for heterocedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, 7, 271-300.

Oja, H. (1981): "On location, scale, skewness and kurtosis of univariate distributions," *Scandinavian Journal of Statistics*, 8, 154-168.

Patton, A.J. (2001): "On the importance of skewness and asymetric dependence in stock returns for asset allocation," Working Paper, London School of Econmomics.

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

Risk (1999): "Risk Software Survey," January, 67-80.

RiskMetrics (1995): "RiskMetrics Technical Document," 3rd Edition, JP Morgan.

Ruppert, Sheather, and Wand, M. (1995): "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, 90, 1257-1270.

Smith, R.L. (1984): "Threshold methods for sample extremes," in Thiago de Oliveira ed., *Statistical Extremes and Applications*, Dordrecht: D. Reidel, 621-638.

– – (1987): "Estimating tails of probability distributions," *The Annals of Statistics*, 15, 1174-1207.

Shephard, M. (1996): “Statistical aspects of ARCH and stochastic volatility models,” in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds., *Time Series Models in Econometrics, Finance and Other Fields*, London: Chapman and Hall, 1-67.

Tauchen, G. (2001): “Notes on Financial Econometrics,” *Journal of Econometrics*, 100, 57-64.

Ziegelmann, F. (2002): “Nonparametric estimation of volatility functions: the local exponential estimator,” *Econometric Theory*, 18, 985-991.