SFB 649 Discussion Paper 2010-009

Predicting extreme VaR: Nonparametric quantile regression with refinements from extreme value theory

Julia Schaumburg*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Predicting extreme VaR: Nonparametric quantile regression with refinements from extreme value theory

Julia Schaumburg[*†]

January 30, 2010

Abstract

This paper studies the performance of nonparametric quantile regression as a tool to predict Value at Risk (VaR). The approach is flexible as it requires no assumptions on the form of return distributions. A monotonized double kernel local linear estimator is applied to estimate moderate (1%) conditional quantiles of index return distributions. For extreme (0.1%) quantiles, where particularly few data points are available, we propose to combine nonparametric quantile regression with extreme value theory. The out-of-sample forecasting performance of our methods turns out to be clearly superior to different specifications of the Conditionally Autoregressive VaR (CAViaR) models.

Keywords: Value at Risk, nonparametric quantile regression, risk management, extreme value theory, monotonization, CAViaR

JEL classification: C14, C22, C52, C53

[*]I would like to thank Melanie Schienle, Nikolaus Hautsch and Vladimir Spokoiny for their instructive and helpful comments on this project. The support from the Deutsche Forschungsgemeinschaft via SFB 649 'Ökonomisches Risiko', Humboldt-Universität zu Berlin, is gratefully acknowledged.

[†]Humboldt-Universität zu Berlin, School of Business and Economics, Institute of Statistics and Econometrics, Chair of Econometrics. Address: Spandauer Str. 1, 10178 Berlin, Germany. Email: julia.schaumburg@wiwi.hu-berlin.de.

1 Introduction

Risk management regulations require banks to estimate market risk measures based on quantiles of loss distributions. According to the Market Risk Amendment to the Basel II Capital Accord of 2004, issued by the Bank for International Settlements, Value at Risk (VaR) is to be calculated daily, using a '99th percentile, one-tailed confidence interval'.[1] Banks are free to choose which VaR model they use, but the recent turbulences on financial markets raise the question to what extent conventional VaR models, e.g. based on historical simulation or estimates of variance-covariance matrices of asset returns, are appropriate.

The aim of this paper is to assess the performance of nonparametric quantile regression as a tool for VaR estimation. The approach requires no assumptions on the form of financial return distributions. We show that in terms of out-of-sample forecasting performance, a monotonized double kernel local linear estimator clearly outperforms competing models on the 1% VaR level. Our benchmarks are different specifications of the Conditionally Autoregressive Value at Risk (CAViaR) models of Engle and Manganelli (2004). By refining nonparametric quantile regression methods with extreme value theory (EVT), we are able to model extreme quantiles (0.1%) accurately.

Several studies exist that compare the forecast performances of different VaR models. See, among others, Kuester et al. (2006), Manganelli and Engle (2001) and Nieto and Ruiz (2008). They take into account a broad variety of models, but nonparametric quantile regression as a tool for VaR estimation is rarely considered. One reason might be the fact that due to regulatory requirements and internal risk management purposes, quantiles associated with low probabilities such as $p = 0.01$ and below are of particular practical interest. For fully nonparametric models, however, the number of data points available to estimate tail quantiles is often not sufficient. Cai and Wang (2008) suggest to estimate VaR and Expected Shortfall using a new nonparametric VaR estimator, combining the Weighted Nadaraya Watson (WNW) estimator of Cai (2002) and the Double Kernel Local Linear (DKLL) estimator of Yu and Jones (1998). In the empirical application, however, only 5% quantile curves are estimated

[1] Amendment to the Capital Accord to incorporate market risks, paragraph B.4(b).

and no forecasts are computed. Chen and Tang (2005) investigate nonparametric VaR estimation, when no regressors are present. Taylor (2008) proposes to combine double kernel quantile regression with exponential smoothing of the dependent variable in the time domain. 1% and 99% VaRs are predicted from the model along with some benchmarks, but extreme quantiles are not considered. Within the framework presented here, nonparametric regression can be utilized to estimate VaR on any probability level of interest.

In recent years, computing power has increased substantially. Thus, estimating nonparametric quantile models induces only little higher computation costs compared to parametric models. But the gain in flexibility is substantial, because nonparametric estimates can also be used as benchmarks to parametric models, which may help to reduce model risk. Therefore, we argue that, also in practice, nonparametric quantile regression should be considered as a serious alternative to common VaR estimation approaches.

Generally, the price to be paid for the flexibility of nonparametric models is slower convergence than in the case of parametric regression. Therefore, when using nonparametric methods to estimate VaR, one major concern is data sparseness in the tails of the return distribution. We address this issue by combining three techniques.

Firstly, the nonparametric technique of double smoothing is applied, i.e. estimation of a distribution by localizing both regressor- and dependent variable observations in order to ease distortions arising from data sparseness. Two candidate double kernel estimators are available, which have similar asymptotic properties. On the basis of a small simulation study, the Double Kernel Local Linear (DKLL) estimator is chosen for the VaR application.

However, investigating in-sample fits obtained from the DKLL estimator, reveals that despite the double smoothing, not all distortions can be eliminated, so that curves that should be smoothly increasing are nonmonotone and have bumps. Therefore, secondly, we adapt the method of monotonization by rearrangement, which has been proposed by Chernozhukov et al. (2009a). To the best of our knowledge, implementing this method is new to the VaR literature. The theoretical finding, that rearrangement weakly improves nonmonotone estimates of monotone functions, is confirmed by our VaR estimation and forecasting results.

Thirdly, since data sparseness is even more severe in case of extreme quantiles, e.g. corresponding to a probability of $p = 0.001$, we apply EVT to estimate quantiles of the standardized nonparametric quantile residuals. The estimation performance of the model on our data set is promising. In a small simulation study, we confirm that the procedure leads to accurate VaR estimates.
The remainder of the paper is structured as follows. Section 2 briefly describes the basic setup of conditional quantile models, including CAViaR models. A new CAViaR specification is introduced, which is directly motivated from the GARCH literature. Section 3.1 outlines the two candidate double kernel estimators. Their finite sample fits are compared via simulation in 3.2. Section 3.3 contains the modelling idea for extreme quantiles, combining nonparametric quantile regression and EVT. The investigated data sets and the backtesting method are summarized in section 4. The empirical results on 1% and 0.1% VaR of four time series of index returns are summarized in section 5. Section 6 concludes.

2 Quantile regression approaches to VaR estimation

2.1 Conditional quantiles

Let $\{Y_t\}_{t=1}^n$ be a strictly stationary time series of portfolio returns and let $\mathbf{X}_t$ be a d-dimensional vector of regressors. The pth conditional quantile of Y_t, denoted by $q_p(\mathbf{x})$, is defined as

$$q_p(\mathbf{x}) = \inf\{y \in \mathbb{R} : F(y|\mathbf{x}) \geq p\} \equiv F^{-1}(p|\mathbf{x}), \tag{2.1}$$

or, equivalently, as the argument that solves

$$\min_{q(\mathbf{X}_t)} \mathbf{E}\big[\big(p - I(Y_t < q(\mathbf{X}_t))\big)\big(Y_t - q(\mathbf{X}_t)\big)|\mathbf{X}_t = \mathbf{x}\big]. \tag{2.2}$$

Both formulations are widely used in the literature. In the seminal paper by Koenker and Bassett (1978) a sample equivalent of (2.2) where $q(\mathbf{X}_t) = \mathbf{X}_t'\boldsymbol{\beta}$, also including the special case $\mathbf{X}_t = 1$, is established. $\boldsymbol{\beta}$ is a vector of unknown parameters and has dimension $d + 1$. The linear quantile model is extended to conditionally heteroskedastic processes in Koenker and Zhao (1996). In En-

gle and Manganelli (2004) conditionally autoregressive quantile functions are estimated using (2.2) with $q(\mathbf{X}_t)$ possibly being nonlinear in parameters, see section 2.2 for some examples. In a number of papers, localized kernel versions of (2.2) are estimated, leading to a nonparametric fit: Yu and Jones (1997) compare the goodness of fit of local constant and local linear models. A varying coefficients and a partially varying coefficients approach are covered in Cai and Xu (2008). On the other hand, Cai (2002), Yu and Jones (1998), Cai and Wang (2008) propose nonparametric methods to estimate the distribution function in (2.1), which, in a second step, is inverted. Section 3.1 contains more details on the three approaches. Wu et al. (2007) model (2.1) without regressors, and Chernozhukov and Umantsev (2001) operationalize a linear version of (2.1).

Following the convention of expressing VaR as a positive number, it is defined as

$$VaR_p^t(\cdot) = -q_p^t(\cdot),$$

where q_p^t is the quantile of the return distribution corresponding to probability p, at time t. VaR_p^t denotes a generic VaR measure which may depend on x and/or a vector of parameters $\boldsymbol{\beta}$. To simplify notation, index t is suppressed in contexts where it does not cause confusion.

2.2 Conditionally autoregressive VaR (CAViaR) Models

The class of Conditional Autoregressive Value at Risk (CAViaR) models, first introduced by Engle and Manganelli (2004), is used to benchmark the forecast performance of the nonparametric VaR estimators considered here. Several comparison studies have done so, for example Kuester et al. (2006) or Taylor (2008). CAViaR models are dynamic VaR models describing the quantile of a random variable at time t, e.g. the return on a financial portfolio, as possibly nonlinear function of its own lags and, in addition, of a vector of observable variables, $\mathbf{X}_t$:

$$VaR_p^t(\boldsymbol{\beta}, \mathbf{X}_t) = \beta_0 + \sum_{i=1}^{r_1} \beta_i VaR_p^{t-i}(\boldsymbol{\beta}, \mathbf{X}_{t-i}) + \sum_{j=1}^{r_2} \beta_j f(\mathbf{X}_{t-j}),$$

where $r = r_1 + r_2 + 1$ is the dimension of $\boldsymbol{\beta}$, the parameter vector that solves

$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{t=1}^{n} \big[p - I\big(Y_t < -VaR_p^t(\boldsymbol{\beta}, \boldsymbol{X}_t)\big)\big](Y_t + VaR_p^t(\boldsymbol{\beta}, \mathbf{X}_t)). \tag{2.3}$$

To simplify notation, the $\mathbf{X}_t$ in parentheses will be dropped in the following. A straightforward choice for $\mathbf{X}_t$ is lagged returns. Following the original article, the specifications used here include the first lagged value of $VaR_p(\cdot)$ and the first lagged value of Y_t, therefore $\mathbf{X}_t = Y_{t-1}$.

Well-known stylized facts on asset returns are, firstly, that they exhibit volatility clustering. It carries over to VaR: if high variation is observed in returns of the recent past, it is likely to continue, and risk is therefore high as well. Secondly, quantiles (or volatility) might react differently according to the sign of past returns. This possibility is captured by the Asymmetric Slope specification

$$VaR_p^t(\boldsymbol{\beta}) = \beta_1 + \beta_2 VaR_p^{t-1}(\boldsymbol{\beta}) + \beta_3 (Y_{t-1})^+ + \beta_4 (Y_{t-1})^-, \tag{2.4}$$

where $(x)^+ = \max(x, 0)$ and $(x)^- = -\min(x, 0)$, but not by the Indirect GARCH(1,1) specification

$$VaR_p^t(\boldsymbol{\beta}) = \sqrt{\beta_1 + \beta_2 (VaR_p^{t-1})^2(\boldsymbol{\beta}) + \beta_3 Y_{t-1}^2}. \tag{2.5}$$

On the other hand, the Asymmetric Slope CAViaR imposes a piecewise linear structure on VaR, although the true functional form might be nonlinear. As pointed out in Kuester et al. (2006), financial returns may also have an autoregressive mean, which is neglected by the above CAViaR specifications. For these reasons we introduce a new specification, called Indirect Autoregressive Threshold GARCH (AR-TGARCH(1,1)) CAViaR:

$$VaR_p^t(\boldsymbol{\beta}) = \beta_1 Y_{t-1} + \big(\beta_2 + \beta_3 (VaR_p^{t-1})^2(\boldsymbol{\beta}) + \beta_4 Y_{t-1}^2 + \beta_5 (Y_{t-1})^2 I(Y_{t-1} < 0)\big)^{\frac{1}{2}}, \tag{2.6}$$

Including the AR term introduces the possibility for a nonzero autoregressive mean, asymmetry is present if $\beta_5 \neq 0$ and the square root allows for a nonlinear functional form.

3 Nonparametric VaR models

3.1 Modelling 1% VaR

In general, estimating nonparametric models requires large amounts of data. Since VaR corresponds to a quantile at the tail of the return distribution, suitable nonparametric quantile estimators should be able to deal with areas where data are sparse. Therefore, from the variety of nonparametric quantile estimators, the Double Kernel Local Linear (DKLL) estimator of Yu and Jones (1998) and the Weighted Double Kernel Local Linear (WDKLL) estimator introduced by Cai and Wang (2008) are considered for the VaR application, because they localize the data in both x- and y-direction, which leads to smoother estimates. For more details, regularity assumptions and asymptotic properties, see the original articles by Cai and Wang (2008) and Yu and Jones (1998).
For notational convenience, observations $\{(X_t, Y_t)\}_{t=1}^n$ are assumed to be drawn from underlying bivariate distribution $F(x, y)$ with density $f(x, y)$. The extension to the multivariate case is straightforward, but requires more tedious notation. Both estimators are defined as inverses of conditional distribution functions as in (2.1). Throughout this section, quantiles of return distributions are discussed, so that VaR corresponds to the negative quantile.
A generic nonparametric method of estimating a conditional distribution $F(y|x)$ is

$$\check{F}(y|x) = \sum_{t=1}^{n} w_t(x) I(Y_t \leq y), \tag{3.1}$$

where $I(A)$ denotes the indicator on the set A and the weights $w_t(x)$ are positive and sum up to one. Choosing equal weights $w = 1/n$ yields the empirical distribution function. Using instead a kernel function with bandwidth parameter h, in the following sometimes abbreviated by $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$, which is often chosen to be a symmetric probability density function, results in the Nadaraya Watson estimator for conditional distribution

$$\check{F}_{NW}(y|x) = \sum_{t=1}^{n} \underbrace{\frac{K_h(x - X_t)}{\sum_{t=1}^{n} K_h(x - X_t)}}_{w_t(x)} I(Y_t \leq y)$$

see for example Li and Racine (2007). It attaches a smooth set of weights to the data, and is known to be monotone increasing and bounded between zero and one. However, it suffers from boundary distortion, as shown by Fan and Gijbels (1996). They advocate the use of local polynomial estimators, the simplest of which is the local linear estimator.

One way to reduce distortions that arise due to a limited number of observations is to smooth not only the observations of the regressor variable X_t, but also the observations of the dependent variable Y_t. This requires the introduction of a second symmetric kernel $W_{h_2}(\cdot)$. Its kernel distribution, which is defined by

$$\int_{-\infty}^{y} W_{h_2}(Y_t - u)du = \Omega\left(\frac{y - Y_t}{h_2}\right), \tag{3.2}$$

with $h_2 < h_1$, can be viewed as a smooth, differentiable version of the indicator function.

In case of the DKLL estimator, as a next step, the conditional distribution value of y is approximated by a linear Taylor expansion around x. The estimate $\tilde{F}(y|x) = \hat{\beta}_0$ is obtained from

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{n} \left(\Omega\left(\frac{y - Y_t}{h_2}\right) - \beta_0 - \beta_1(X_t - x)\right)^2 K_{h_1}(x - X_t). \tag{3.3}$$

Solving for $\hat{\beta}_0$ yields the explicit expression for the conditional distribution function estimator,

$$\tilde{F}(y|x) = \sum_{t=1}^{n} \underbrace{\frac{K_{h_1}(x - X_t)\left[S_2 - (x - X_t)S_1\right]}{\sum_{t=1}^{n} K_{h_1}(x - X_t)\left[S_2 - (x - X_t)S_1\right]}}_{w_t(x)} \Omega\left(\frac{y - Y_t}{h_2}\right), \tag{3.4}$$

where

$$S_l = \sum_{i=1}^{n} K\left(\frac{x - X_t}{h_1}\right)(x - X_i)^l, \qquad l = 1, 2.$$

(3.4) is a version of (3.1) where the kernel distribution function $\Omega(\cdot)$ in (3.2) replaces the indicator. The DKLL quantile estimator $\tilde{q}_p(x)$, the sample analogue to (2.1), is then defined by

$$\tilde{q}_p(x) = \inf\left\{y \in \Re : \tilde{F}(y|x) \geq p\right\} \equiv \tilde{F}^{-1}(p|x). \tag{3.5}$$

with $\tilde{F}$ from (3.4). In finite samples, $\tilde{F}(y|x)$ might not always be monotonically increasing. In such cases, however, the inverse is not defined. Yu and Jones (1998) suggest the following implementation scheme: For $\tilde{q}_{1/2}(x)$, any value satisfying (3.5) is chosen; for $p > 1/2$, the largest, and for $p < 1/2$, the smallest solutions to (3.5) are taken as quantile estimates.
In this paper, a stronger procedure is applied, avoiding to delete estimated values. Chernozhukov et al. (2009a) show that any nonmonotone estimate of a monotone function can be improved in terms of common metrics, such as the L_p-norm, by rearranging. For the case of a monotone increasing (decreasing) function, the point estimates are sorted in ascending (descending) order. Making use of the results derived in Chernozhukov et al. (2009a), nonmonotone distribution estimates are rearranged before inverting. We will utilize this useful method more extensively in section 5.2 to monotonize conditional VaR curves. In the present context of monotonizing the estimated distribution function, a further effect is that quantile crossing is circumvented, as pointed out in Chernozhukov et al. (2009b). Estimated values greater than one are discarded.
The Weighted Double Kernel Local Linear (WDKLL) of Cai and Wang (2008) estimator is a combination of the DKLL estimator and the Weighted Nadaraya Watson (WNW) estimator of Cai (2002). The indicator in (3.1) is replaced by distribution function (3.2). Additionally, in order to avoid boundary distortions known to occur for standard Nadaraya Watson type estimators, a set of weight functions $p_t(x)$ is multiplied to the kernel values. The weight functions depend on the data $X_1, ..., X_n$ and on locations x. Here they are chosen to fulfill the discrete moment conditions of the simplest local polynomial estimator, the local linear, which are

$$\sum_{t=1}^{n} p_t(x) = 1 \quad \text{and} \quad \sum_{t=1}^{n} p_t(x)(X_t - x)K_h(X_t - x) = 0. \tag{3.6}$$

Fan and Gijbels (1996) show that as a consequence of these conditions, design dependent local polynomial estimators automatically adjust at the boundary of the support of x, and that they can adapt to different designs.
Functions $p_t(x)$ fulfilling (3.6) are not unique. One possibility to identify them is to use the idea underlying empirical likelihood: The product, or equivalently

the sum of the logarithms of all $p_t(x)$ is maximized subject to the constraints (3.6). The corresponding (Lagrange) objective function is

$$L = \sum_{t=1}^{n} log\left[p_t(x)\right] + \lambda_1 \left(\sum_{t=1}^{n} p_t(x) - 1\right) + \lambda_2 \left(\sum_{n=1}^{n} (X_t - x) p_t(x) K_h(x - X_t)\right). \tag{3.7}$$

By taking derivatives and solving the first order conditions, $p_t(x;\lambda)$ can be derived as

$$p_t(x;\lambda) = n^{-1}\left[1 + \lambda(X_t - x)K_h(x - X_t)\right] \tag{3.8}$$

Plugging (3.8) back into the objective function gives

$$L = \frac{1}{nh}\sum_{t=1}^{n} log\left[1 + \lambda(X_t - x)K_h(x - X_t\right], \tag{3.9}$$

which is maximized by finding the root of $L'(\lambda) = 0$ numerically, e.g. by Newton's Method. The obtained parameter λ_0 is used in (3.8), which gives the unique weights.
Putting everything together, the WDKLL conditional distribution estimator is defined by

$$\hat{F}\left(y|x\right) = \sum_{t=1}^{n} \frac{p_t(x)K_{h_1}\left(X_t - x\right)}{\sum_{t=1}^{n} p_t(x)K_{h_1}\left(X_t - x\right)} \Omega\left(\frac{y - Y_t}{h_2}\right), \tag{3.10}$$

and the corresponding WDKLL estimate of the pth conditional quantile function is

$$\hat{q}_p(x) = \inf\{y \in \Re : \hat{F}(y|x) \geq p\} \equiv \hat{F}^{-1}(p|x) \tag{3.11}$$

It always exists because $\hat{F}(y|x)$ is, by construction, between zero and one and monotone in y (see Cai (2002)). In Cai and Wang (2008) it is shown that both conditional distribution and quantile estimators are design adaptive, a feature that is not shared by ordinary Nadaraya Watson type estimators. In particular, no boundary correction is necessary. The replacement of the indicator entails additional smoothness, especially at the outer regions of the support.

3.2 Comparing DKLL and WDKLL estimator

From a theoretical point of view, the WDKLL estimator is slightly superior to the DKLL estimator, because it ensures monotonicity and does not require rearrangement. Furthermore, it is explicitly set up for time series data. Still, since the aim of this paper is to find the estimator which is best suited for estimation of VaR, i.e. tail quantiles, a small simulation is carried out comparing DKLL and WDKLL estimators. An ARCH(1) process with starting value 0 is generated according to

$$Y_t = -0.4X_t + \underbrace{\sqrt{0.4(1+X_t^2)}}_{\sigma(X_t)}\epsilon_t, \tag{3.12}$$

where $X_t = Y_{t-1}$, and the error term $\epsilon_t \sim iidN(0,1)$. Conditional quantiles are estimated using both WDKLL ($\hat{q}(x)$) and DKLL ($\tilde{q}(x)$) estimators for three different sample sizes, $n = 200$, $n = 500$ and $n = 1000$, conditional on two values of X_t, $x = -0.75$ and $x = 1.25$. For both local constant fit of the WDKLL estimator and local linear fit of the DKLL estimator, the Gaussian kernel is used, while the uniform kernel is used for smoothing the dependent variable.

Table 3.1 contains results on estimates of the Integrated Square Error (ISE)

$x = -0.75$

	5% ISE	25% ISE	Median ISE	75% ISE	95% ISE
WDKLL	0.009	0.012	0.015	0.018	0.024
DKLL	0.014	0.018	0.021	0.024	0.032

$x = 1.25$

	5% ISE	25% ISE	Median ISE	75% ISE	95% ISE
WDKLL	0.033	0.048	0.059	0.074	0.100
DKLL	0.025	0.035	0.043	0.053	0.070

Table 3.1: Quantiles of $\widehat{ISE}$ for $n = 1000$.

as goodness of fit measure, which is defined for some function $f(x)$ and an estimate $\hat{f}(x)$, as

$$ISE = \int \left(\hat{f}(x) - f(x)\right)^2 dx.$$

Here, ISE measures the squared distance between estimate and true quan-

tile curve. It can be estimated by discretizing the integral using the trapezoid formula

$$\widehat{ISE} = \sum_{i=2}^{m-1} (\hat{f}(x^i) - f(x^i))^2 \Delta + \frac{1}{2}(\hat{f}(x^1) - f(x^1))^2 \Delta + \frac{1}{2}(\hat{f}(x^m) - f(x^m))^2 \Delta, \quad (3.13)$$

where $x^1 < ... < x^m$ is a grid of x-values and $\Delta = x^i - x^{i-1}$ is the same for all $i = 2, ...m$. Here, median, 5%, 25%, 75% and 95% quantiles of $\widehat{ISE}$ are computed. All results are derived using 500 replications. Interestingly, for the grid point relatively close to the process mean, $x = -0.75$, the WDKLL estimator has smaller $\widehat{ISE}$, but for $= 1.25$, where less data are available, the fit of the DKLL estimator is slightly better. As the differences are small, both estimators are considered to be well suited for our application. However, based on the simulation result we choose the DKLL estimator for the estimation of index return VaR in section 5.2.

3.3 Modelling 0.1% VaR

For extreme quantiles, usually very few data points are available, so that fully nonparametric regression does not yield reliable estimates. Extreme value theory (EVT) is an alternative to model extreme quantiles. In the following a method of incorporating extreme value theory into CAViaR models, which was introduced by Manganelli and Engle (2001), is adapted to obtain VaR estimates for $p = 0.001$ from a nonparametric model.
The strategy is to first calculate the standardized quantile residuals,

$$\frac{\hat{\epsilon}_\theta^t}{\hat{q}_\theta^t} = \frac{Y_t - \hat{q}_\theta^t}{\hat{q}_\theta^t} = \frac{Y_t}{\hat{q}_\theta^t} - 1.$$

Under the assumption that the implemented model is correct, they should be i.i.d., which is a necessary condition for applying standard extreme value estimators. McNeil and Frey (2000) employ a similar technique to estimate VaR for $p = 0.01$ from a GARCH residual series. p denotes the (very low) probability of interest, and θ corresponds to a moderately low probability for which the quantile can be estimated nonparametrically, for example $\theta = 0.01$ or $\theta = 0.05$. Reformulating the definition of the pth quantile of portfolio returns in terms of

the θth quantile yields

$$\begin{aligned} P[Y_t < q_p^t] &= P\left[Y_t < q_\theta^t - q_\theta^t + q_p^t\right] \\ &= P\left[\frac{Y_t}{q_\theta^t} - 1 > \frac{q_p^t}{q_\theta^t} - 1\right] = p. \end{aligned}$$

The inequality sign is switched assuming that q_p^t is a negative number. Let

$$\frac{q_p^t}{q_\theta^t} - 1 \equiv z_p$$

denote the pth quantile of the standardized residuals. Finally, the pth quantile of the original time series of portfolio returns can be estimated with EVT:

$$\frac{\hat{q}_p^t}{\hat{q}_\theta^t} - 1 = \hat{z}_p \quad \Leftrightarrow \quad \hat{q}_p^t = \hat{q}_\theta^t(\hat{z}_p + 1). \tag{3.14}$$

with EVT methods, and some rearranging.
Again, $\widehat{VaR}_p^t = -\hat{q}_p^t$. In the remainder of this section, the basic idea of the standard peaks over threshold (POT) method, which is used to obtain $\hat{z}_p$ in (3.14), is described very briefly, following Embrechts et al. (1997).
Large observations which exceed a high threshold can be approximated reasonably well by the generalized Pareto distribution with distribution function

$$G_{\xi,\beta}(x) = \begin{cases} -(1 + \xi x/\beta)^{1/\xi} & \text{for } \xi \neq 0 \\ 1 - e^{x/\beta} & \text{for } \xi = 0 \end{cases} \tag{3.15}$$

with shape parameter ξ and scale parameter $\beta > 0$. The support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\frac{\beta}{\xi}$ if $\xi < 0$. The parameters can be consistently estimated if the threshold exceedances are independent, regardless of the true underlying distribution, see Smith (1987). In general, given a high threshold u and a random variable Y, the probability of Y exceeding u at most by x is given by

$$F_u(x) = P[Y - u \leq x | Y > u] = \frac{F(x+u) - F(u)}{1 - F(u)}. \tag{3.16}$$

Balkema and de Haan (1974) and Pickands (1975) show that for a large class of distribution functions F it is possible to find a positive function $\beta(u)$ such that

$$\lim_{u \to y_0} \sup_{0 \leq x < y_0 - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \tag{3.17}$$

with y_0 corresponding to the right endpoint of F. Rearranging (3.16) and using $F_u(\cdot) \approx G_{\xi,\beta}(\cdot)$, it holds that

$$1 - F(u + x) \approx [1 - F(u)][1 - G_{\xi,\beta}(x)].$$

Then, $1 - G_{\xi,\beta}(x)$ can be obtained by estimating the GPD parameters by maximum likelihood. Let N_u denote the number of exceedances over threshold u. A common way of estimating $S(u) := 1 - F(u)$ is to use the empirical distribution function $\frac{N_u}{n}$. Subsituting the estimates,

$$\widehat{S(u + x)} = \frac{N_u}{n} \left(1 + \hat{\xi} \left(\frac{x}{\hat{\beta}} \right) \right)^{-\frac{1}{\hat{\xi}}}. \tag{3.18}$$

The quantile can be estimated by inverting (3.18), employing a change of variables $y = u + x$ and fixing the distribution value at the probability of interest: $F(y) = p$. Therefore, the quantile estimator $\hat{q}_p$ is obtained from

$$\begin{aligned} 1 - p &= \frac{N_u}{n} \left(1 + \hat{\xi} \left(\frac{y - u}{\hat{\beta}} \right) \right)^{-\frac{1}{\hat{\xi}}} \\ \Leftrightarrow \quad \hat{q}_p &= u + \left[\left((1 - p) \frac{N_u}{n} \right)^{-\hat{\xi}} - 1 \right] \cdot \frac{\hat{\beta}}{\hat{\xi}}. \end{aligned} \tag{3.19}$$

4 Data and backtesting method

We analyze four data sets of daily index returns. The longest available time series of each are used to compute in-sample fits, leaving out 1000 observations for out-of-sample forecasting. Table 4 summarizes the data. The end date of the in-sample period is 04/05/2004.

Realizations of quantiles cannot be observed. Therefore, backtesting of the models is carried out using the dynamic quantile (DQ) out-of-sample test developed in Engle and Manganelli (2004) to test and compare the performance

	DAX	FTSE 100	EuroSTOXX	S&P 500
start date	01/05/1965	01/03/1984	01/02/1987	06/26/1969
no. of observations	9954	4999	4216	8787
mean	0.0167	0.0259	0.0205	0.0246
median	0	0.0195	0.0567	0.0019
0.5% quantile	-4.0200	-3.4698	-4.9243	-3.003
99.5% quantile	3.6530	3.2573	4.1567	3.2247
skewness	-0.41	-0.79	-0.33	-1.41
kurtosis	11.28	13.57	8.41	38.21

Table 4.1: Data summary. All returns in percent.

of CAViaR models. Define the hit function

$$Hit_t \equiv I(Y_t < -VaR_p^t) - p \tag{4.1}$$

which equals $-p$ if the return is below the forecasted quantile and $(1-p)$ if VaR is exceeded. If the chosen model is correct,

1. $\mathbf{E}[Hit_t|\Omega_t] = 0$, where Ω_t is any information known at t, and consequently,
2. Hit_t is uncorrelated with its own lags and
3. $P(Y_t < -VaR_p^t) = p$, i.e. the unconditional probability of VaR exceedance equals p.

Thus, VaR is estimated correctly, if for each day independently, the probability of exceeding it equals p. For the DQ test, a regression equation

$$Hit_t = \mathbf{X}_t'\boldsymbol{\theta} + u_t, \quad u_t = \begin{cases} -p & \text{with prob. } 1-p \\ 1-p & \text{with prob. } p \end{cases} \tag{4.2}$$

is estimated, where $\mathbf{X}_t$ is an r-dimensional vector containing any variables potentially correlated with Hit_t. The null hypothesis

$$\mathbf{H}_0 : \theta_1 = ... = \theta_r = 0$$

can be tested by a Wald test for joint significance. Therefore, the test statistic is

$$DQ = \frac{1}{n} \cdot \frac{\mathbf{Hit}'\mathbf{X}[\mathbf{X}'\mathbf{X}]\mathbf{X}'\mathbf{Hit}}{p(1-p)} \sim \chi^2_r \text{ as } n, n_{IS} \to \infty,$$

where n denotes the number of out-of sample forecasts, n_{IS} is the number of observations used for estimating the model, and **Hit** and $\mathbf{X}$ are the vectors containing observations of the dependent variable and the regressor matrix, respectively.
Following, for example, Engle and Manganelli (2004), Kuester et al. (2006) and Taylor (2008), the information set consists of a constant, four lagged values of Hit_t and the respective estimate of VaR_p^{t-1}.

5 Application to stock index returns

5.1 1% CAViaR

For estimating the parameters of the CAViaR models, an algorithm similar to the one proposed in the original paper is applied, see Engle and Manganelli (2004). A grid search is conducted by generating a large number of random vectors, the dimension of which corresponds to the number of model parameters. The five vectors which lead to the lowest values of the objective function (2.3) are selected and fed into a simplex optimization algorithm. The final parameter vector is chosen to be the one minimizing (2.3).
Table 5.1 reports the results on the evaluation of the Asymmetric Slope, GARCH and AR-TGARCH CAViaR models. They perform similarly. All in-sample coverages, i.e. the shares of VaR exceedances in the estimation period, are very close to 1%. This is not surprising since the objective function ensures that the parameters are chosen in this way. Except for the DAX, out-of-sample coverages are in an acceptable range as well. However, almost all p-values of the DQ test are close to zero, i.e. the null hypothesis of independent VaR exceedances has to be rejected on common significance levels in almost all cases.
Comparing forecast accuracy, it turns out that the the results obtained from the new AR-TGARCH CAViaR model are similar to the results from the Asymmetric Slope model. However, the news impact curves shown in figure 5.1, i.e. the

DAX			
	Asymmetric Slope	GARCH	AR-TGARCH
in-sample (%)	0.995	1.025	0.975
out-of sample (%)	5.0	5.3	05.1
DQ p-value	0.0	0.0	0.0

FTSE 100			
	Asymmetric Slope	GARCH	AR-TGARCH
in-sample (%)	1.020	1.000	0.980
out-of sample (%)	1.0	0.8	1.1
DQ p-value	0.0	0.0	0.0

EuroSTOXX 50			
	Asymmetric Slope	GARCH	AR-TGARCH
in-sample (%)	0.997	1.020	0.996
out-of sample (%)	1.8	1.7	1.9
DQ p-value	0.000061	0.021	0.000067

S&P 500			
	Asymmetric Slope	GARCH	AR-TGARCH
in-sample (%)	1.002	1.013	1.001
out-of sample (%)	0.6	0.7	0.6
DQ p-value	0.000029	0.0022	0.000027

Table 5.1: DQ test results for CAViaR models as well as in-sample and out-of sample share (coverage) of VaR exceedances (in percent).

reactions of VaR to different magnitudes of the lagged return, reveal that the new specification resembles the nonparametric VaR estimate better than the other two models. Nevertheless, from the DQ test results it must be concluded that none of the considered CAViaR specifications captures the dynamics underlying the VaR processes.

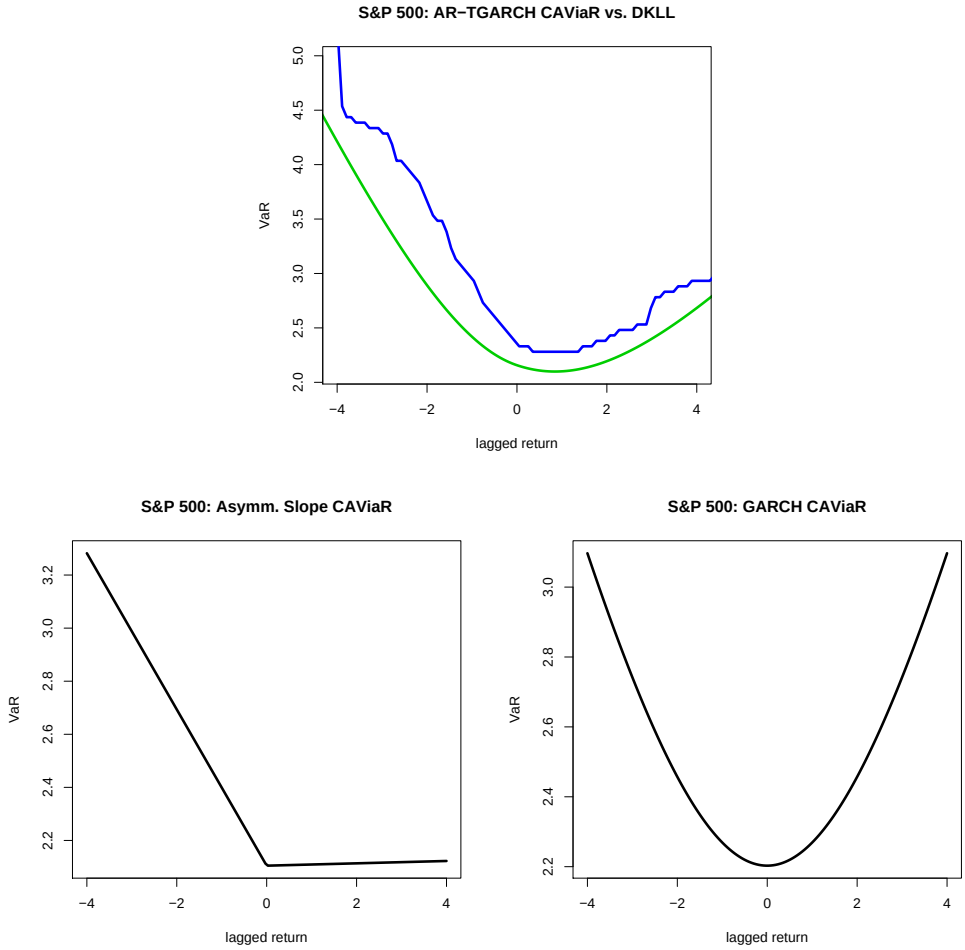

Figure 5.1: News impact curves of TGARCH(1,1) CAViaR (green) together with DKLL estimate (blue), of Asymmetric Slope and GARCH(1,1) CAViaR, for S&P 500

5.2 Nonparametric 1% VaR

When forecasting from a nonparametric model, one has to balance two effects occuring at the boundary areas: The support from which predictions of the dependent variable can be computed is limited to the range in which the estimated function is located. This means that for outlying lagged returns, which are not in the support of the estimated curve, no forecasts for VaR exist. On the other hand, often only few data points are available at boundary areas, so that outliers have more influence and the resulting curve may show distortions. Therefore, one has to decide carefully about the range of the grid at which the

function is evaluated, balancing possible distortions against a limited range of regressor values to compute forecasts from.
For forecasting 1% quantiles of a conditional distribution, the DKLL estimator is used due to its double smoothing property, which eases distortions and leads to smoother quantile curves. Furthermore, in our simulation, it obtained a slightly superior finite sample fit compared to the WDKLL estimator (see section 3.1).
However, the performance of the DKLL estimator can be improved even further by making use of the monotonization method proposed by Chernozhukov et al. (2009a). Whenever curves are not monotonically decreasing on the left of the minimum and monotonically increasing on the right, estimated values are rearranged in descending and ascending order, respectively. Chernozhukov et al. (2009a) shows that this procedure of rearranging point estimates weakly reduces the estimation error for any nonmonotone estimate of a monotone function. To illustrate possible changes in the in-sample fit, figures 5.2 and 5.3

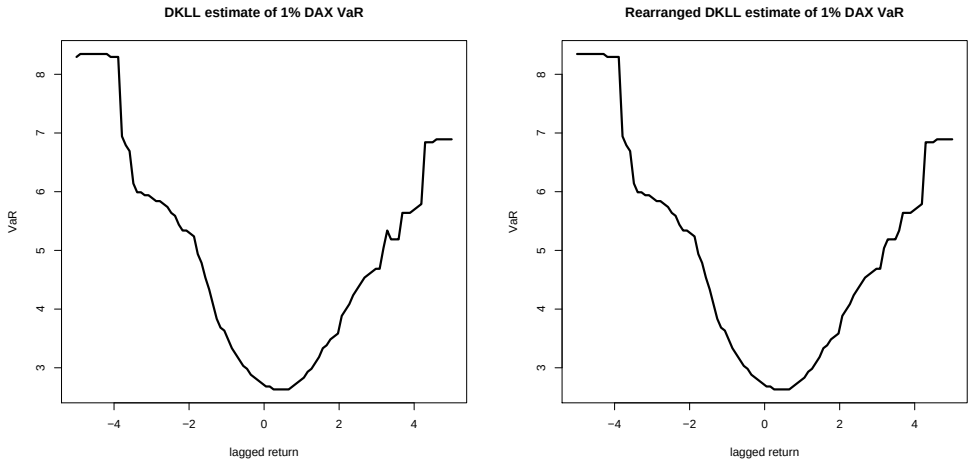

Figure 5.2: Original and rearranged DKLL estimates of 1% conditional DAX VaR curve

show the original as well as the rearranged 1% VaR curves of DAX and Eurostoxx. All curves cover 99% of the data.
The backtesting results of original DKLL and rearranged DKLL estimates are summarized in table 5.2. It reports in-sample and out-of-sample coverages, i.e. the shares of VaR exceedances in the estimation and forecasting periods, respectively, as well as the p-value of the out-of-sample DQ test described in

subsection 4. Whenever values in the columns are different, they are superior for the rearranged estimates: In-sample and out-of-sample coverages are closer to 1% in case of the FTSE return series. Furthermore, the DQ test p-value is higher, indicating that the null hypothesis of independent hits is further away from rejection than for the original DKLL model. For EuroSTOXX, in-sample coverage is closer to 1% as well. The results for DAX and S&P 500, on the other hand, were not affected by the rearrangement. Therefore, our estimation results confirm the findings of Chernozhukov et al. (2009a).

The conclusions drawn from backtesting the nonparametric model are simi-

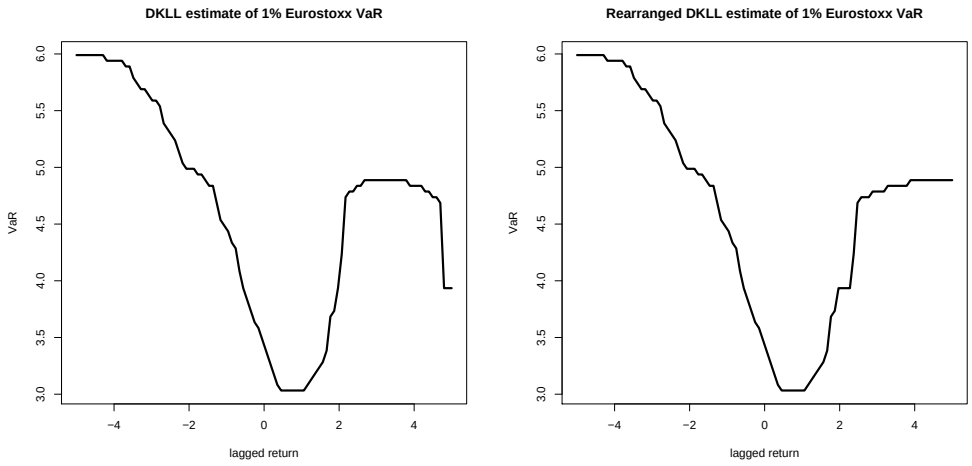

Figure 5.3: Original and rearranged DKLL estimates of 1% conditional Eurostoxx VaR curve

lar for FTSE, EuroSTOXX and S&P 500. Table 5.2 reveals that there seems to be a slight tendency to overestimate VaR, so that exceedances stay below the expected amount. The DQ test p-values, on the other had, indicate that no severe clustering of exceedances is present. The large number of observations justifies the choice of a test significance level of 1%, so that the null hypothesis of independent hits is not rejected even for S&P 500.

The picture is different in the case of DAX VaR. Table 5.3 contains a direct comparison of results obtained from the rearranged DKLL model and, due to their similarity, from only two of the CAViaR specifications. The two CAViaR models clearly underestimate VaR in the forecasting period, and the null hypothesis of independent hits has to be rejected on any significance level. The latter is true for the rearranged DKLL model, too, although the p-value is slightly

	DAX		FTSE	
	DKLL orig.	DKLL rearr.	DKLL orig.	DKLL rearr.
in-sample (%)	0.804	0.804	0.900	0.940
out-of-sample (%)	1.2	1.2	0.4	0.5
DQ p-value	0.00014	0.00014	0.10	0.48

	EuroSTOXX		S&P500	
	DKLL orig.	DKLL rearr.	DKLL orig.	DKLL rearr.
in-sample (%)	0.807	0.830	0.945	0.945
out-of-sample (%)	0.5	0.5	0.3	0.3
DQ p-value	0.53	0.53	0.012	0.012

Table 5.2: DQ test results for original and rearranged DKLL models as well as in-sample and out-of sample share (coverage) of VaR exceedances (in percent).

greater than zero. However, the out-of-sample coverage obtained by the model is close to 1%. Thus, the model fits obtained from the rearranged DKLL estimator are clearly superior to the results from the different CAViaR models.

	DKLL rearr.	A.S. CAV.	GARCH CAV.
in-sample (%)	0.804	0.995	1.025
out-of-sample (%)	1.2	5.0	5.3
DQ p-value	0.00014	0.0	0.0

Table 5.3: DAX: DQ test results as well as in-sample and out-of sample share (coverage) of VaR exceedances (in percent).

5.3 Nonparametric EVT-augmented 0.1% VaR estimates

Following the procedure described in section 3.3, standardized residuals are computed from the rearranged DKLL estimate and the time-varying 0.1% quantile of time series Y_t is calculated according to (3.14). For completeness, the results from the EVT-augmented Asymmetric Slope- and GARCH CAViaR specifications are computed as well, following Manganelli and Engle (2001).

Table 5.4 contains both in-sample and out-of-sample share of DAX-VaR exceedances for the four considered models. The choice of DAX is motivated by the fact that forecasts of DAX VaR in the investigated time period seems to be

	DKLL & EVT	DKLL	EVT-A.S.-CAV.	EVT-GARCH-CAV.
in-sample (%)	0.13	0.09	0.10	0.16
out-of sample (%)	0.1	0	1.4	1.9

Table 5.4: In-sample and out-of sample shares (coverages) of VaR exceedances for 0.1% VaR (in percent).

particularly challenging, as it turned out in section 5.2. Only the DKLL estimator achieved the correct out-of-sample coverage, but the DQ test indicated that it failed to produce a time series of independent VaR exceedances.
Table 5.4 shows that the model performances are similar to the results on 1% VaR forecasting: The CAViaR models underestimate VaR. This finding is not surprising, as for computation of the standardized residuals the quantile residuals from the 1% VaR model are used. Therefore, the goodness of fit of the model corresponding to the 'moderate' probability carries over to the extreme quantile.
On the other hand, the fractions of VaR exceedances are very close to the underlying probabilities for both pure and EVT-augmented DKLL estimates. This similarity of results is surprising. One would expect more stable results for the EVT-augmented estimates, because only few observations are available in the extreme tails. Due to the shortness of the time horizon, no DQ test p-values are reported.[2]
In order to assess whether the EVT-augmented DKLL model leads to an improved forecast performance over the basic DKLL model, a small simulation is done where 60000 observations are generated from the ARCH(1) model

$$Y_t = 0.1Y_{t-1} + \sqrt{10^{-7} + 0.3\epsilon_{t-1}^2} \cdot \epsilon_t, \quad \epsilon_t \sim t(4).$$

Each model is estimated using the first 10000 observations. Two forecast horizons N=20000 and N=50000 are considered. Table 5.5 contains the results on coverage and DQ test p-values. In case of the shorter forecasting period both models perform similarly, but the p-value obtained for the extended forecast horizon clearly indicates that the EVT-augmented DKLL estimator describes the extreme conditional VaR more accurately than the basic DKLL estimator.

[2]In case of the DKLL estimate, no exceedances are achieved, so that there is no variation in one column of the regressor matrix, and the p-value is not defined.

	N=20000		N=50000	
	DKLL & EVT	DKLL	DKLL & EVT	DKLL
In-sample	0.11	0.1	0.08	0.1
Out-of sample	0.115	0.125	0.08	0.12
DQ p-value	0.0000088	0.000002	0.85	$7.3 \cdot 10^{-12}$

Table 5.5: Backtesting results: 0.1% VaR forecasts for simulated time series (in percent). N corresponds to the number of forecast periods.

Thus, it can be concluded that the combination of standardized nonparametric residuals and extreme value theory forms a solid alternative to estimate extreme VaRs. The method is therefore a valuable complement to the rearranged DKLL estimator which we suggest to use for quantiles corresponding to moderately low probabilities such as $p = 0.01$.

6 Conclusion

In this paper, we propose a way to nonparametrically estimate conditional Value at Risk that is associated with very small probabilities such as $p = 0.01$ and $p = 0.001$. A rearranged Double Kernel Local Linear VaR estimator as well as a version of the latter augmented by extreme value theory are investigated and applied to index return time series. Forecasts are benchmarked against the widely used CAViaR models. In terms of generating a conditionally independent sequence of VaR exceedances over the forecasting period, the performance of all considered CAViaR models is poor, while the rearranged DKLL estimates performs well. Furthermore, refining nonparametric quantile regression by extreme value theory yields promising results.

References

Balkema, A. A. and L. de Haan (1974). Residual life time at great age. *The Annals of Probability 2*, 792–804.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory 18*, 169–192.

Cai, Z. and X. Wang (2008). Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics 147*, 120–130.

Cai, Z. and X. Xu (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association 103*, 1595–1608.

Chen, S. X. and C. Y. Tang (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics 3*, 227–255.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009a). Improving point and interval estimates of monotone functions by rearrangement. *Biometrika 96*, 559–575.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009b). Quantile and probability curves without crossing. *Econometrica (forthcoming)*.

Chernozhukov, V. and L. Umantsev (2001). Conditional value-at-risk: Aspects of modelling and estimation. *Empirical Economics 26*, 271–292.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events*. Springer.

Engle, R. F. and S. Manganelli (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*, 367–381.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability 66. Chapman & Hall.

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

Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for arch models. *Econometric Theory 12*, 793–813.

Kuester, K., S. Mittnik, and M. S. Paolella (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics 4*, 53–89.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics*. Princeton University Press.

Manganelli, S. and R. F. Engle (2001). Value at risk models in finance. *ECB Working Paper Series Working Paper No. 75*.

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

Nieto, M. R. and E. Ruiz (2008). Measuring financial risk: Comparison of alternative procedures to measure var and es. *Universidad Carlos III de Madrid Working Paper No. 08-73*.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics 3*, 119–131.

Smith, R. L. (1987). Estimating the tails of probability distributions. *The Annals of Statistics 15*, 1174–1207.

Taylor, J. W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics 6*, 382–406.

Wu, W. B., K. Yu, and G. Mitra (2007). Kernel conditional quantile estimation for stationary processes with application to value at risk. *Journal of Financial Econometrics*, 1–18.

Yu, K. and M. Jones (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis 25*, 159–166.

Yu, K. and M. C. Jones (1998). Local linear quantile regression. *Journal of the American Statistical Association 93*, 228–237.

SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Volatility Investing with Variance Swaps" by Wolfgang Karl Härdle and Elena Silyakova, January 2010.
002 "Partial Linear Quantile Regression and Bootstrap Confidence Bands" by Wolfgang Karl Härdle, Ya'acov Ritov and Song Song, January 2010.
003 "Uniform confidence bands for pricing kernels" by Wolfgang Karl Härdle, Yarema Okhrin and Weining Wang, January 2010.
004 "Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model" by Nikolaus Hautsch and Fuyu Yang, January 2010.
005 "The Impact of Macroeconomic News on Quote Adjustments, Noise, and Informational Volatility" by Nikolaus Hautsch, Dieter Hess and David Veredas, January 2010.
006 "Bayesian Estimation and Model Selection in the Generalised Stochastic Unit Root Model" by Fuyu Yang and Roberto Leon-Gonzalez, January 2010.
007 "Two-sided Certification: The market for Rating Agencies" by Erik R. Fasten and Dirk Hofmann, January 2010.
008 "Characterising Equilibrium Selection in Global Games with Strategic Complementarities" by Christian Basteck, Tijmen R. Daniels and Frank Heinemann, January 2010.
009 "Predicting extreme VaR: Nonparametric quantile regression with refinements from extreme value theory" by Julia Schaumburg, February 2010.