Predicting extreme Value at Risk: Nonparametric quantile regression with refinements from extreme value theory

Julia Schaumburg*

December 9, 2011

Abstract

A framework is introduced allowing to apply nonparametric quantile regression to Value at Risk (VaR) prediction at any probability level of interest. A monotonized double kernel local linear estimator is used to estimate moderate (1%) conditional quantiles of index return distributions. For extreme (0.1%) quantiles, nonparametric quantile regression is combined with extreme value theory. The abilities of the proposed estimators to capture market risk are investigated in a VaR prediction study with empirical and simulated data. Possibly due to its flexibility, the out-of-sample forecasting performance of the new model turns to be superior to competing models.

Keywords: Value at Risk, nonparametric quantile regression, risk management, extreme value statistical applications, monotonization

*Humboldt-Universität zu Berlin, School of Business and Economics, Institute of Statistics and Econometrics, Chair of Econometrics, Spandauer Str. 1, 10178 Berlin, Germany, email: julia.schaumburg@wiwi.hu-berlin.de, phone +49-(0)30-2093-5603, fax +49-(0)30-2093-.

1 Introduction

In bank regulation, the effectiveness of capital requirements in preventing funding shortfalls rests upon the estimation accuracy of market risk measures, the most widely used of which is Value at Risk (VaR). According to the Market Risk Amendment to the Basel II Capital Accord of 2004, issued by the Bank for International Settlements, VaR is to be calculated daily, using a '99th percentile, one-tailed confidence interval'.

Not only against the background of the financial turbulences during the crisis 2007-2009, there is a practical need for VaR models that are rich enough to capture the dynamics of quickly changing market environments, while being parsimonious at the same time in order to avoid overfitting. In this context, kernel-based quantile regression is a natural choice. Being fully nonparametric, the approach does not require any structural assumptions, neither on the form of the VaR function nor on the shape of the loss distribution. Furthermore, the primary interest is to predict quantiles accurately, not to estimate structural parameters, so that we do not lose interpretability by employing nonparametric estimation methods.

Until now, however, nonparametric quantile regression has not been widely used in risk management practice. One possible reason is that precisely because no structure is imposed on the regression function, the speed of convergence is slower than in parametric regression settings, and generally more observations are needed for accurate estimation. Therefore, when considering nonparametric quantile regression for VaR prediction, one has to address the problem of data sparseness at the tails of the loss distribution, which gets more severe when considering quantiles corresponding to extreme probabilities.

This paper proposes a framework allowing to operationalize nonparametric quantile regression as a VaR estimation tool for both moderate and extreme probabilities. We include only past returns as regressor, which is sufficient, as a maximum of information can be exploited and any kinds of nonlinearities are captured within the modelling approach. In addition, the model is simple, overfitting is avoided and problems arising from limited observations are minimized. A double kernel-type estimator is used to smooth distortions. Additionally, monotonization by rearrangement is applied at the boundary of the estimated function. Finally, in order to predict VaR corresponding to extreme probability levels (such as 0.1%), the peaks over threshold method is incorporated and applied to the standardized nonparametric quantile residuals, resulting in an estimator that combines nonparametric quantile regression and extreme value theory (EVT).

Several studies have compared 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 a broad variety of models into account, but nonparametric quantile regression as a tool for VaR estimation is rarely included. There are three exceptions we are aware of: 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 section, however, only 5% quantile curves are estimated 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. Although the Basel Committee on Banking Supervision asks banks to report VaR for a 10-day holding period, we focus on one-day-ahead forecasts, which is in line with the literature.

In contrast to the studies already available in the literature, we present a method that allows to nonparametrically estimate VaR corresponding to any probability that might be of practical interest. We estimate, predict and backtest 1% and 0.1% VaRs for four sets of index returns and a simulated time series. Our focus is on gains to loosening assumptions in comparison to existing VaR models. Therefore, in the empirical application, we choose to benchmark our model against the most flexible parametric VaR models, the Conditional Autoregressive Value at Risk (CAViaR) models of Engle and Manganelli (2004). They allow to set up different linear and nonlinear specifications, including the lagged VaR estimate as a regressor, and they do not rely on distributional assumptions.

We find that although the CAViaR models obtain almost perfect in-sample fits in the case of 1% VaR, the Double Kernel Local Linear (DKLL) estimator of Yu and Jones (1998) often outperforms them in terms of out-of-sample backtesting results, especially when the estimation period is short relative to the forecasting period. Results are promising for 0.1%

VaR as well. The superiority of the EVT-refined DKLL estimator over the plain DKLL estimator is shown in a small simulation study. It turns out that, especially when the estimation window is small relative to the forecasting period, the extreme value theory-refined nonparametric model predicts extreme VaRs very accurately.

Section 2 outlines the basic setup of conditional quantile models, before describing CAViaR models. The Double Kernel Local Linear estimator used in the following is presented in Section 3. Furthermore, the incorporation of extreme value theory into the model is explained. The investigated data sets and the backtesting method are summarized in Section 4. The empirical results on 1% and 0.1% VaR estimation are stated in Section 5. In Section 6, the performance of the EVT-refined nonparametric model is further assessed via a small simulation study. Section 7 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

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

where $I(A)$ denotes the indicator function on some set A. 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 d-dimensional vector of unknown parameters. The linear quantile model is extended to conditionally heteroskedastic processes in Koenker and Zhao (1996). In Engle and Manganelli (2004) conditional 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. 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 these 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 $\mathbf{x}$ 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 Conditional 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, $\mathbf{X}_t$:

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

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

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

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 (AR) mean, which is neglected by the CAViaR specifications, which artificially set the mean return equal to zero. For these reasons we combine the positive features of the above, by allowing for nonlinearity and asymmetric effects of past returns, and additionally incorporate an AR mean, by introducing an alternative specification, called Indirect Autoregressive Threshold GARCH (AR-TGARCH(1,1)) CAViaR:

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

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.

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). Our new AR-TGARCH specification fits into this procedure.

3 Nonparametric quantile regression with refinements from extreme value theory

3.1 Double Kernel Local Linear VaR regression

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) is chosen for the VaR application, because it 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 article by Yu and Jones (1998). The Weighted Double Kernel Local Linear estimator of Cai and Wang (2008), a Nadaraya Watson type estimator, forms an alternative to the DKLL estimator. In a small simulation study, which is not reported here, the DKLL estimator performed slightly better at design points at the boundary of the support of the data. Therefore, we chose it for our application.

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. The estimator is defined as inverse of a conditional distribution function 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 $I(\cdot)$ is again an indicator function 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 of a conditional distribution

$$\check{F}_{NW}(y|x) = \sum_{t=1}^{n} \underbrace{\frac{K_h(x - X_t)}{\sum_{t=1}^{n} K_h(x - X_t)}}_{w_t(x)} I(Y_t \leq y), \tag{3.2}$$

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

$$\int_{-\infty}^{y} W_{h_2}(Y_t - u)du = \Omega\left(\frac{y - Y_t}{h_2}\right), \tag{3.3}$$

with h_2 being the bandwidth parameter, can be viewed as a smooth, differentiable version of the indicator function.

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

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{t=1}^{n} \left(\Omega\left(\frac{y - Y_t}{h_2}\right) - \beta_0 - \beta_1(X_t - x)\right)^2 K_{h_1}(x - X_t), \tag{3.4}$$

where $h_1 > h_2$. Solving for $\hat{\beta}_0$ yields the explicit expression for the conditional distribution function estimator,

$$\tilde{F}(y|x) = \sum_{t=1}^{n} \underbrace{\frac{K_{h_1}(x - X_t)[S_2 - (x - X_t)S_1]}{\sum_{t=1}^{n} K_{h_1}(x - X_t)[S_2 - (x - X_t)S_1]}}_{w_t(x)} \Omega\left(\frac{y - Y_t}{h_2}\right), \tag{3.5}$$

where

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

(3.5) is a version of (3.1) where the kernel distribution function $\Omega(\cdot)$ in (3.3) replaces the indicator. The DKLL quantile estimator $\tilde{q}_p(x)$, the sample analogue to (2.1), is then defined by

$$\tilde{q}_p(x) = \inf\left\{y \in \Re : \tilde{F}(y|x) \geq p\right\} \equiv \tilde{F}^{-1}(p|x). \tag{3.6}$$

with $\tilde{F}$ from (3.5). 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 (3.6) is chosen; for $p > 1/2$, the largest, and for $p < 1/2$, the smallest solutions to (3.6) are taken as quantile estimates.

In this paper, a stronger procedure is applied, avoiding to delete estimated values. Chernozhukov et al. (2009) show that any nonmonotone estimate of a monotone function can be improved in terms of common metrics, such as the L_p-norm, by simple rearranging. In an earlier work, Dette et al. (2006) propose a similar method of smoothed rearrangements. For the case of a monotone increasing (decreasing) function, the point estimates are sorted in ascending (descending) order. Making use of these theoretical results, nonmonotone distribution estimates are rearranged before inverting. In the present context of monotonizing the estimated distribution function, a further effect is that quantile crossing is circumvented. Estimated values greater than one are discarded.

3.2 Refining nonparametric quantile regression with extreme value theory

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 0.1% VaR estimates from a nonparametric model.

The strategy is to first calculate the standardized quantile residuals,

$$\frac{\hat{\epsilon}^t_{p_2}}{\hat{q}^t_{p_2}} = \frac{Y_t - \hat{q}^t_{p_2}}{\hat{q}^t_{p_2}} = \frac{Y_t}{\hat{q}^t_{p_2}} - 1. \tag{3.7}$$

p denotes the (very low) probability of interest, and p_2 corresponds to a moderately low probability for which the quantile can be estimated nonparametrically, for example $p_2 = 0.01$ or $p_2 = 0.05$. McNeil and Frey (2000) employ a similar technique to estimate 1% VaR from a GARCH residual series. An EVT-augmented nonparametric kernel distribution estimator is also considered by MacDonald et al. (2011), who show consistency of their method via Bayesian inference.

Reformulating the definition of the pth quantile of portfolio returns in terms of the p_2th quantile yields

$$\begin{aligned} P[Y_t < q^t_p] &= P\left[Y_t < q^t_{p_2} - q^t_{p_2} + q^t_p\right] \\ &= P\left[\frac{Y_t}{q^t_{p_2}} - 1 > \frac{q^t_p}{q^t_{p_2}} - 1\right] = p. \end{aligned}$$

The inequality sign is switched assuming that $q^t_{p_2}$ is a negative number. Let

$$z_p \equiv \frac{q^t_p}{q^t_{p_2}} - 1$$

denote the $(1-p)$th quantile of the standardized residuals. It is then estimated by the peaks over threshold (POT) method, though it can be estimated by other methods, such as the Hill estimator, as well. A number of applications employ the POT method to forecast extreme quantiles; for a selection of applications and an investigation of its finite sample properties see El-Arouia and Diebolt (2002). An estimate for the pth return quantile can be expressed by means of $\hat{z}_p$ and $\hat{q}^t_{p_2}$:

$$\frac{\hat{q}^t_p}{\hat{q}^t_{p_2}} - 1 = \hat{z}_p \quad \Leftrightarrow \quad \hat{q}^t_p = \hat{q}^t_{p_2}(\hat{z}_p + 1). \tag{3.8}$$

Again, $\widehat{VaR}^t_p = -\hat{q}^t_p$. In the remainder of this section, the underlying extreme value arguments, which is used to obtain $\hat{z}_p$ in (3.8), is discussed very briefly, following Embrechts et al. (1997).

Large observations which exceed a high threshold can be approximated reasonably well by the Generalized Pareto Distribution (GPD) with distribution function

$$G_{\xi,\beta}(x) = \begin{cases} -(1+\xi x/\beta)^{1/\xi} & \text{for } \xi \neq 0 \\ 1 - e^{x/\beta} & \text{for } \xi = 0 \end{cases} \tag{3.9}$$

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

$$F_u(x) = P[Y - u \leq x | Y > u] = \frac{F(x+u) - F(u)}{1 - F(u)}. \tag{3.10}$$

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

$$\lim_{u \to y_0} \sup_{0 \leq x < y_0 - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0, \tag{3.11}$$

with y_0 corresponding to the right endpoint of F. Rearranging (3.10) 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 $\frac{N_u}{n}$. Substituting the estimates,

$$\widehat{S(u+x)} = \frac{N_u}{n} \left(1 + \hat{\xi} \left(\frac{x}{\hat{\beta}} \right) \right)^{-\frac{1}{\hat{\xi}}}. \tag{3.12}$$

The quantile can be estimated by inverting (3.12), 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

$$1 - p = \frac{N_u}{n}\left(1 + \hat{\xi}\left(\frac{y-u}{\hat{\beta}}\right)\right)^{-\frac{1}{\hat{\xi}}}$$
$$\Leftrightarrow \quad \hat{q}_p = u + \left[\left((1-p)\frac{n}{N_u}\right)^{-\hat{\xi}} - 1\right] \cdot \frac{\hat{\beta}}{\hat{\xi}}. \tag{3.13}$$

In general, extreme value methods require the underlying data to be i.i.d. Although computing standardized residuals in (3.7) should remove most of the dynamics, one cannot eliminate the possiblitity of remaining autocorrelation. However, under some conditions on the dependence structure (see e.g. Drees (2003) for details), the relationship between the limiting distributions of the maxima of a dependent but strictly stationary sequence, $(Y_t)_{t\in\mathbb{N}}$ say, and a white noise sequence $(\widetilde{Y}_t)_{t\in\mathbb{N}}$ with the same distribution function F may be described by the so-called extremal index $\theta \in (0,1]$. If the distribution of normalized threshold exceedances in the sequence $(\widetilde{Y}_t)$ converges to an extreme value distribution $G(x)$, as in (3.11), then it can be shown that the equally normalized exceedances of $(Y_t)_{t\in\mathbb{N}}$ converge to $G^\theta(x)$, see Embrechts et al. (1997). Thus, the same limiting extreme value distribution may be used, while changing only the normalization parameters.

Intuitively, if our sequence of standardized residuals possesses an extremal index which is < 1, then its extremal behavior is asymptotically the same as that of a shorter white noise sequence with the same distribution. However, it might still be interesting to find out about the extent of deviation from white noise for a given data set. The extremal index may be estimated by the so-called Runs Method where θ is computed as the conditional probability that a threshold exceedance is followed by a run of r non-exceedances. The idea is that periods in between clusters of exceedances are longer than periods between independent exceedances (for details see Embrechts et al. (1997), Chapter 8). The higher the clustering tendency, the fewer runs will be present. The estimator is

$$\widehat{\theta} = \frac{\sum_{t=1}^{n-r} I(A_t)}{\sum_{t=1}^{n} I(Y_t > u)}. \tag{3.14}$$

where $I(\cdot)$ is again the indicator function and

$$A_t = \{Y_t > u, Y_{t+1} \leq u, ..., Y_{t+r} \leq u\} \tag{3.15}$$

Drees (2003) shows that a wide variety of financial time series, including ARMA and ARCH processes, may be estimated by the POT maximum likelihood estimator which we use here. He finds that the only difference to i.i.d. data is an increased variance of the quantile estimator, but this drawback does not affect our results, as our goal is forecasting accuracy, which is checked via backtesting.

4 Data and backtesting method

We analyze four data sets of daily index returns, DAX, FTSE 100 (FTSE), EuroSTOXX 50 (EuroSTOXX) and S&P 500 (S&P). The longest available time series of each are used to compute in-sample fits. The common end date of the in-sample period is 28/02/2003. We predict VaR for the subsequent 1000 days, until the end of 2006 (29/12/2006). As a second step, we take the same forecast period, but additionally include 300 days to check whether model performances worsen when the data contains the beginning of the financial crisis. The 1300-day forecast period ends on 22/02/2008. Table 4 summarizes some features of the data.

	DAX	FTSE	EuroSTOXX	S&P
start date	03/10/1966	03/01/1984	02/01/1987	31/07/1970
end in-sample period	28/02/2003	28/02/2003	28/02/2003	28/02/2003
no. of observations	9500	4998	4215	8500
mean	0.020	0.026	0.021	0.028
median	0.000	0.020	0.057	0.007
0.5% quantile	-4.107	-3.470	-4.924	-3.009
99.5% quantile	3.686	3.258	4.157	3.239
skewness	-0.422	-0.794	-0.326	-1.468
kurtosis	11.158	13.571	8.414	39.075

Table 4.1: Data summary for the time series used in the long estimation period.

The reason for using a rather long estimation period relative to the forecast horizon, is that our aim is to assess and compare the quality of the nonparametric model in capturing market risk, using as much information as possible. We are aware that in real life risk management, available time series are typically much shorter. Therefore, in Section 5.2.2, we additionally estimate our models using only the last 1000 data points, from 30/04/1999 to 28/02/2003 and forecast 1% VaRs for the subsequent 200 and 1000 days.

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 of VaR models. From the variety of alternatives, we choose this particular test, firstly because it is the standard test to compare CAViaR and other models, and secondly, because it is based on a moment condition rather than an augmented regression, where VaR excess indicators are modelled as function of some information set. Escanciano and Olmo (2010) show that many popular VaR backtests do not take estimation risk into account, which may lead to incorrect limiting distributions of the test statistic. However, in the context of Theorem 2 in their paper, they also point out that the out-of-sample DQ test does not suffer from this drawback.
Define the binary variable

$$Hit_t \equiv I(Y_t < -VaR_p^t) - p.$$

If the chosen model is correct,

$$\mathbf{E}[Hit_t|\Omega_t] = 0, \tag{4.1}$$

where Ω_t is any information known up to time t. Thus, VaR is estimated correctly, if independently for each day of the forecasting period, the probability of exceeding it equals p. Note that this also implies that Hit_t is uncorrelated with its own lagged values. Let Z_t denote a K-vector of variables potentially related to Hit_t, and let **Z** denote the $(N \times K)$-matrix stacking the values of Z_t, where N is the number of observations in the forecast period. Then, the moment condition in (4.1) can be checked using the test statistic

$$DQ = \frac{\mathbf{Hit}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Hit}}{p(1-p)}, \tag{4.2}$$

where **Hit** is a vector containing all values of Hit_t. Under the null hypothesis that (4.1) holds, DQ is asymptotically χ^2 distributed with K degrees of freedom. In analogy to Engle and Manganelli (2004), Kuester et al. (2006) and Taylor (2008), we include a constant, four lagged values of Hit_t and the current VaR estimate in the information set.
It can be seen immediately that for our information set, the DQ test statistic requires at least one out-of-sample VaR exceedance in order to be defined. Otherwise, the lagged

values of Hit_t would cause multicollinearity in the matrix $\mathbf{Z}$. However, when considering extreme VaR, as we do in Section 5.3, it might well be that there are no exceedances within the forecasting period (note that the correct number of exceedances for 1000 forecasts of 0.1% VaR is 1). Therefore, in order to be able to compare the different models while taking the possibility of no exceedances into account, we employ the test proposed by Kupiec (1995) which is an unconditional test of the correctness of the achieved share of VaR exceedances. Define the indicator variable

$$I_t \equiv I(Y_t < -VaR_p^t).$$

The idea of the Kupiec test is to check whether $\mathbf{E}[I_t] = p$, in which case the number of exceedances

$$ex_N = \sum_{t=n+1}^{n+N} I_t$$

has a binomial distribution with parameters N and p. Under the null hypothesis of correct coverage, the corresponding likelihood ratio statistic

$$LR_{Kup} = 2\log\left[\left(1 - \frac{ex_N}{N}\right)^{N-ex_N}\left(\frac{ex_N}{N}\right)^{ex_N}\right] - 2\log\left[(1-p)^{N-ex_N}p^{ex_N}\right].$$

is asymptotically χ^2 distributed with one degree of freedom.

5 Application to stock index returns

5.1 Monotonicity of quantile curves

When forecasting from a nonparametric model, one has to balance two effects occurring 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.

We estimate the time-varying conditional 1% VaRs of DAX, EuroSTOXX, S&P and FTSE using the DKLL estimator. Due to its double smoothing property, distortions are eased and quantile curves are smoother. Additionally, we make use of the monotonization method proposed by Chernozhukov et al. (2009). 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. To illustrate possible changes in the in-sample fit, Figure 5.1 shows the original as well as the rearranged 1% VaR curves of DAX and EuroSTOXX. Both curves cover 99% of the data.

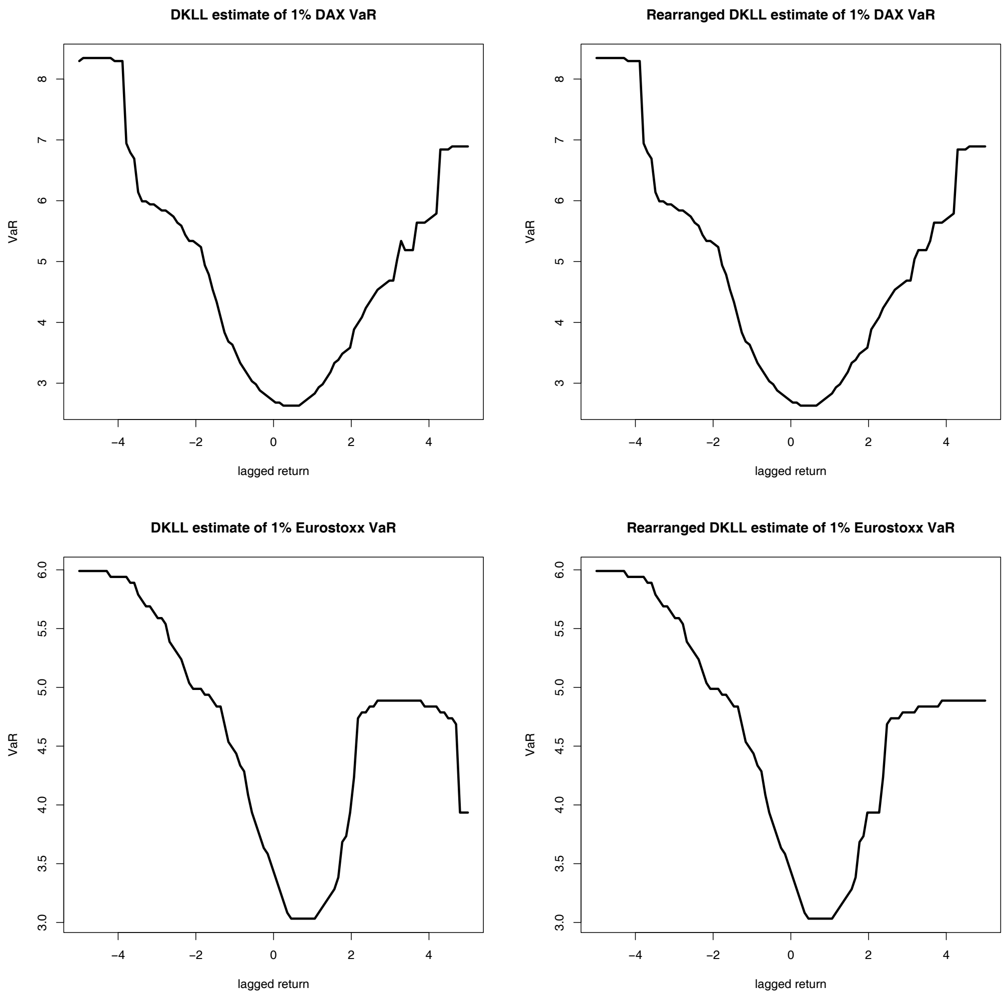

Figure 5.1: Original and rearranged DKLL estimates of 1% conditional DAX VaR curve (upper panel) EuroSTOXX VaR curve (lower panel).

Table 5.1 compares backtesting results on original and rearranged DKLL fits, on a forecast horizon of 1000 days. In-sample and out-of-sample coverages are aimed to be as close as possible to the underlying probability, in this case 1%. The P-value of the LR test described in Section 4 expresses the highest significance level at which the variables in the information set are jointly significant. Therefore, a larger P-value indicates that the null hypothesis of independent VaR exceedances is more likely not to be rejected, suggesting that a model is more adequate

	DAX		FTSE	
	DKLL orig.	DKLL rearr.	DKLL orig.	DKLL rearr.
in-sample (%)	0.78	0.78	0.94	0.95
out-of-sample (%)	1.00	1.00	0.40	0.50
DQ P-value	0.182	0.182	0.614	0.830

	EuroSTOXX		S&P500	
	DKLL orig.	DKLL rearr.	DKLL orig.	DKLL rearr.
in-sample (%)	0.81	0.81	0.97	0.97
out-of-sample (%)	0.50	0.50	0.30	0.30
DQ P-value	0.859	0.859	0.544	0.544

Table 5.1: DQ test results for original and rearranged DKLL models as well as in-sample and out-of sample share of VaR exceedances (in percent). The forecast period is 1000 observations.

The theoretical results from Chernozhukov et al. (2009), that rearranging weakly improves estimation, are confirmed by our empirical results: 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 LR test P-value substantially increases, indicating that the null hypothesis of the LR test is 'further away' from rejection than in the case of the original DKLL model. Whenever we mention results for the DKLL estimator in the following, it refers to the rearranged version.

5.2 Comparing 1% VaR predictions

5.2.1 Long estimation period

Table 5.2 lists backtest results of the CAViaR models and the rearranged DKLL estimates which are obtained from the large data set data set described in Table 4. Generally, the in-sample exceedance shares achieved by all three CAViaR specifications are very close to the underlying probability 1%. In contrast, the DKLL estimator has a slight tendency to overestimate VaR, leading to in-sample coverages below 1%. The news impact curves shown in Figure 5.2 reveal that the AR-TGARCH CAViaR specification resembles the nonparametric VaR estimate better than the other two models.

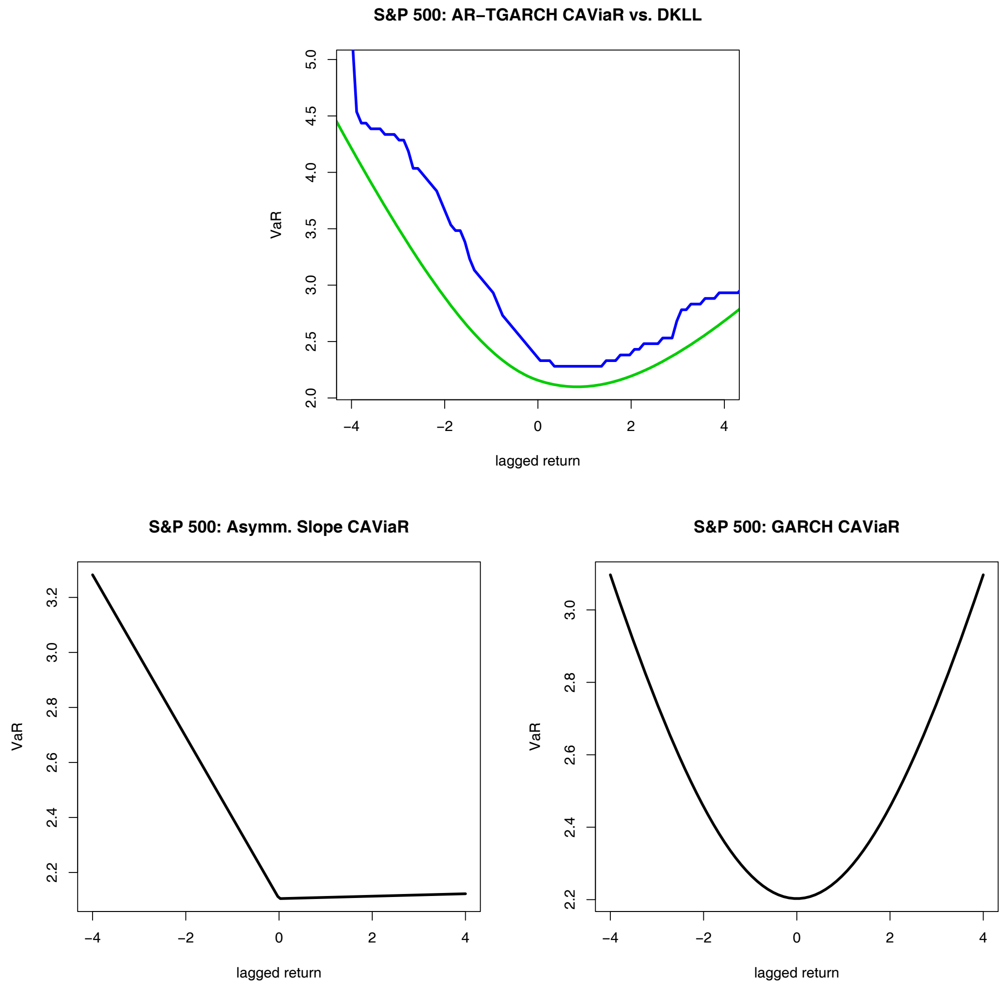

Figure 5.2: News impact curves, i.e. reactions of S&P VaR to different magnitudes of lagged returns ('news'), of TGARCH(1,1) CAViaR (lower curve in the upper panel) together with DKLL estimate (upper curve in the upper panel), of Asymmetric Slope and GARCH(1,1) CAViaR.

In terms of out-of-sample forecasting, on the other hand, results differ among the four indices. In predicting DAX VaR, the CAViaR models perform quiete poorly on both forecast horizons. Out-of-sample coverages are too high, and the DQ test P-values raise doubt that the models are able to generate conditionally independent VaR exceedances and correct coverage. In contrast, the DKLL estimate achieves more accurate out-of-sample exceedance rates in predicting VaR over both 1000 and 1300 days, and the test results suggest that the model is appropriate at least for the shorter forecast horizon.

	Asymm. Slope	GARCH	AR-TGARCH	DKLL
		DAX		
in-sample	1.01	1.01	1.04	0.78
out-of-sample (1000)	1.50	1.50	1.50	1.00
out-of-sample (1300)	1.54	1.54	1.61	1.07
DQ P-value (1000)	0.054*	0.011**	0.054*	0.182
DQ P-value (1300)	0.043**	0.009***	0.039**	0.040**
		FTSE		
in-sample	1.02	1.00	1.00	0.94
out-of-sample (1000)	0.60	0.60	0.60	0.50
out-of-sample (1300)	1.08	1.16	1.08	1.00
DQ P-value (1000)	0.005***	0.007***	0.007***	0.830
DQ P-value (1300)	0.040**	0.104	0.059*	0.011**
		EuroSTOXX		
in-sample	1.04	0.99	0.99	0.81
out-of-sample (1000)	0.70	0.80	0.80	0.50
out-of-sample (1300)	0.76	0.92	0.92	0.62
DQ P-value (1000)	0.970	0.058*	0.057*	0.859
DQ P-value (1300)	0.980	0.015**	0.015**	0.031**
		S&P		
in-sample	1.01	1.00	0.97	0.97
out-of-sample (1000)	0.30	0.30	0.30	0.30
out-of-sample (1300)	0.92	1.00	0.69	1.15
DQ P-value (1000)	0.547	0.285	0.547	0.544
DQ P-value (1300)	0.042**	0.000***	0.079*	0.008***

Table 5.2: Backtesting results for 1% VaR models. In-sample and out-of-sample exceedance probabilities in %. Considered forecast horizons are 1000 and 1300 observations. Models which are rejected by the DQ test are marked with * for rejection on 10%, ** on 5% and *** on 1% significance level.

In the case of FTSE VaR, results are mixed: The DKLL estimator overestimates VaR in the short prediction period (coverage is 0.5%), where, however, it yields a good fit according to the backtest. For 1300 forecasts, it achieves the correct coverage, but the DQ P-value drops sharply. The CAViaR models, on the other hand, are strongly rejected in the short horizon, but show a better performance in the longer one.

The picture for EuroSTOXX is somewhat similar to that of FTSE, except that the results from the CAViaR models now differ more strongly among each other, and the Asymmetric Slope CAViaR beats all the other models considered. The P-value obtained by the

DKLL estimator again drops when moving from the short to the longer forecast horizon, but it is still above the P-values of GARCH and AR-TGARCH CAViaR, which, on the other hand, perform better in terms of out-of-sample coverages.

The results for S&P VaR show a different structure. Although all coverages within the short prediction period are low, the DQ test indicates adequate out-of-sample fits. For the extended horizon, all P-values drop, such that the GARCH CAViaR and the DKLL models are even rejected at a 1% significance level. One possible reason is that the additionally included observations exhibit some dynamics which are not well captured by these models, leading to a clustering of VaR exceedances. Interestingly, the AR-TGARCH is least affected by this effect.

The AR-TGARCH CAViaR model does not outperform the other two CAViaR models and the DKLL model systematically, but its results are less varying: For both in-sample and out-of-sample forecast horizons, its coverage and backtest results are often better than the results of one of the two others. We attribute this finding to the fact that the model combines the features of Asymmetric Slope and Indirect GARCH specification, and it is therefore more universally applicable.

Summing up, the out-of-sample VaR prediction results produced by the fully nonparametric DKLL estimator are satisfactory except for the extended forecast horizon in the case of the S&P. The CAViaR models are strong competitors, however, they have the drawback that it is not possible to detect one parameterization that systematically dominates others. As it is often the case with parametric models, the question remains which one to choose in practical applications. The DKLL model, on the other hand, outperforms at least one of the CAViaR models in most cases, and is therefore the most robust alternative.

5.2.2 Short estimation period

In real life risk management, time series available for the estimation of VaR models are rarely as long as the ones we investigate in the previous section. For this reason, we repeat the estimation using only 1000 observations, i.e. roughly the last four years up to 28/02/2003, and forecast VaRs for both the subsequent 200 and 1000 days. Table 5.3 lists the backtesting results.

	Asymm. Slope	GARCH	AR-TGARCH	DKLL
		DAX		
in-sample	1.10	1.10	0.90	0.80
out-of-sample (200)	2.50	1.50	2.50	0.50
out-of-sample (1000)	0.70	0.80	0.80	0.10
DQ P-value (200)	0.409	0.839	0.410	0.967
DQ P-value (1000)	0.603	0.061*	0.671	0.195
		FTSE		
in-sample	1.00	1.00	1.10	0.70
out-of-sample (200)	1.00	1.00	1.00	0.50
out-of-sample (1000)	0.70	0.60	0.60	0.10
DQ P-value (200)	0.810	1.000	0.975	0.997
DQ P-value (1000)	0.020**	0.008***	0.007***	0.225
		EuroSTOXX		
in-sample	1.10	1.00	1.10	1.10
out-of-sample (200)	1.50	1.00	1.50	1.00
out-of-sample (1000)	0.40	0.40	0.50	0.20
DQ P-value (200)	0.985	1.000	0.992	1.000
DQ P-value (1000)	0.674	0.652	0.713	0.356
		S&P		
in-sample	1.00	1.00	1.10	0.90
out-of-sample (200)	1.00	0.50	1.00	0.50
out-of-sample (1000)	0.20	0.10	0.20	0.10
DQ P-value (200)	0.536	0.742	0.390	0.975
DQ P-value (1000)	0.206	0.110	0.210	0.212

Table 5.3: Backtesting results for 1% VaR models which are estimated using only 1000 observations. In-sample and out-of-sample exceedance probabilities in %. Considered forecast horizons are 200 and 1000 observations. Models which are rejected by the DQ test are marked with * for rejection on 10%, ** on 5% and *** on 1% significance level.

The good performance of the DKLL estimator carries over to the short estimation period. Although VaR estimates are again too conservative in particular for the longer forecasting period, the null hypothesis of valid moment conditions tested by the DQ test is never rejected even on a 10% significance level. The CAViaR models also overestimate VaR for the subsequent 1000 observations, and additionally, they are rejected by the DQ test in the case of FTSE. Based on these results, it can be said that the DKLL estimator is also applicable when the estimation period is rather short, and keeps yielding reliable VaR forecasts even for the distant future.

5.3 Comparing extreme VaR predictions

Following the procedure described in Section 3.2, 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 (3.8). The underlying 'moderate' probability is chosen to be 1%. Similarly, VaR estimates obtained from the EVT-augmented CAViaR models are computed, following Manganelli and Engle (2001). As mentioned at the end of Section

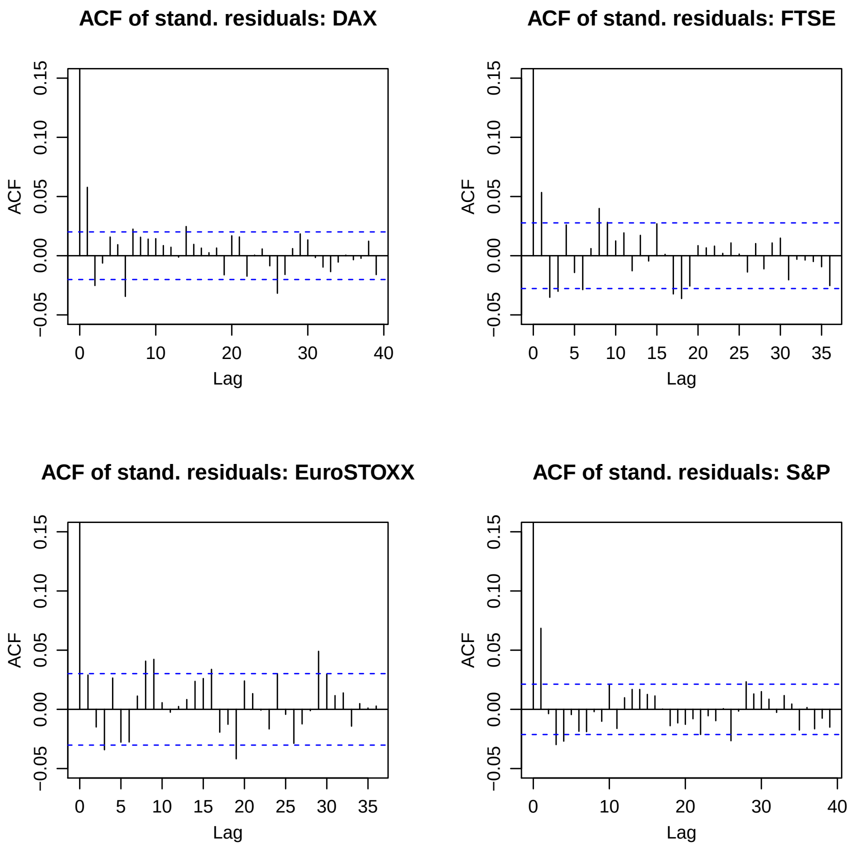

Figure 5.3: Autocorrelation functions of the standardized nonparametric residuals for the four indices. The dashed lines are 95% confidence intervals.

3.2, the data should be checked for dependence before applying extreme value methods. Figure 5.3 shows autocorrelation functions (ACFs) for the standardized residuals from the nonparametric model, together with 95% confidence intervals. Although the magnitude of the autocorrelation is not very high, for some lags, the confidence intervals are exceeded. Therefore, we carry out Ljung-Box tests on independence based on 20 lags, the results of which imply significant autocorrelation on a 1% confidence level for all four considered models and all model specifications. Fitting simple AR models to the standardized residuals, however, removes the autocorrelation entirely, see Table 5.3. The corresponding results for the three CAViaR specifications are very similar, and thus, not shown here. They are available upon request. Since McNeil et al. (2005) state that ARMA processes with innovations drawn from fat-tailed distributions exhibit values of the extremal index $\theta < 1$, we also estimate the extremal indices using the Runs Method (3.14) described in Section 3.2 and report them in the last column of Table 5.3. The parameter r, corresponding to the length of runs, was set to 30 for all three indices, after finding that the estimated $\widehat{\theta}$ was very robust with respect to plausible choices of r. As McNeil et al. (2005) point out, the distribution of the maximum of n dependent data points with extremal index θ can be approximated by the associated i.i.d. series with $n\theta$ observations. Given the large number of data points in all our samples, we conclude that the loss in accuracy due to dependence in the standardized residuals is not too severe, so that we can apply the proposed method to estimate the 0.1%-VaRs.

	LB P-value	LB P-value for AR residuals	AR lag order	$\widehat{\theta}$
DAX	0.000	0.529	7	0.80
FTSE	0.000	0.203	9	0.91
EuroSTOXX	0.001	0.357	9	0.85
S&P	0.000	0.224	4	0.88

Table 5.4: The first column reports the outcomes (P-values) of the Ljung-Box (LB) test on independence of the standardized residuals from the DKLL model. The null hypothesis of independence is always rejected on a 1% confidence level. The second column contains the LB test results after fitting an autoregressive (AR) model to the standardized residuals. In all cases, the null hypothesis cannot be rejected. The selected lag order is reported in the third column. The forth column contains estimates of the extremal index θ.

Table 5.5 contains both in-sample and out-of-sample shares of 0.1% VaR exceedances for the four considered models. Only the long estimation period is considered. However, the simulation study in Section 6 contains a discussion of results from the nonparametric model for extreme quantiles, based on a shorter space of time. Due to the occurrence of no VaR exceedances within the prediction period, we use the Kupiec test instead of the DQ test for backtesting. It checks the correctness of the achieved unconditional coverage via a likelihood ratio approach, which is based on the Bernoulli likelihood, see Section 4. In contrast to the outcomes of the 1% VaR analysis, in-sample VaR exceedance shares

	Asymm. Slope	GARCH	AR-TGARCH	DKLL
	DAX			
in-sample	0.11	0.17	0.09	0.13
out-of-sample(1000)	0.00	0.00	0.00	0.10
out-of-sample(1300)	0.15	0.15	0.15	0.23
Kupiec P-value (1000)	0.157	0.157	0.157	1.000
Kupiec P-value (1300)	0.570	0.570	0.570	0.203
	FTSE			
in-sample	0.14	0.16	0.14	0.16
out-of-sample(1000)	0.10	0.20	0.10	0.10
out-of-sample(1300)	0.15	0.38	0.15	0.15
Kupiec P-value (1000)	1.000	0.379	1.000	1.000
Kupiec P-value (1300)	0.570	0.014**	0.570	0.570
	EuroSTOXX			
in-sample	0.19	0.26	0.31	0.24
out-of-sample(1000)	0.10	0.20	0.20	0.00
out-of-sample(1300)	0.23	0.31	0.31	0.15
Kupiec P-value (1000)	1.000	0.379	0.379	0.157
Kupiec P-value (1300)	0.203	0.058*	0.058*	0.570
	S&P			
in-sample	0.20	0.22	0.19	0.12
out-of-sample(1000)	0.10	0.00	0.00	0.00
out-of-sample(1300)	0.15	0.08	0.08	0.00
Kupiec P-value (1000)	1.000	0.157	0.157	0.157
Kupiec P-value (1300)	0.570	0.784	0.784	0.107

Table 5.5: Backtesting results for 0.1% VaR models. In-sample and out-of-sample exceedance probabilities in %. Considered forecast horizons are 1000 and 1300 observations. Models which are rejected by the DQ test are marked with * for rejection on 10%, ** on 5% and *** on 1% significance level.

achieved by the DKLL estimator are now less conservative, but instead always slightly higher than the target probability 0.1%. On the other hand, out-of sample coverage and backtest results are remarkably good especially for DAX and EuroSTOXX, where it shows best results on one of the two considered forecast horizon, but also for DAX, where its coverages are very close to 0.1%. Concerning S&P, the Asymmetric Slope CAViaR model yields the most accurate fit, except for the in-sample exceedance rate, which is closer to 0.1% in the case of the DKLL model. According to the Kupiec test, the differences to the nominal coverage are rarely significant, the only exception being the GARCH and AR-TGARCH CAViaRs when predicting 1300 days of EuroSTOXX VaR, and GARCH CAViaR for the extended forecast period of FTSE VaR.

It was pointed out in Kuester et al. (2006), that for comparison of VaR prediction strategies, the focus should not be limited to one or two probability levels, but one should take a range of quantiles into account when deciding which model is the best.

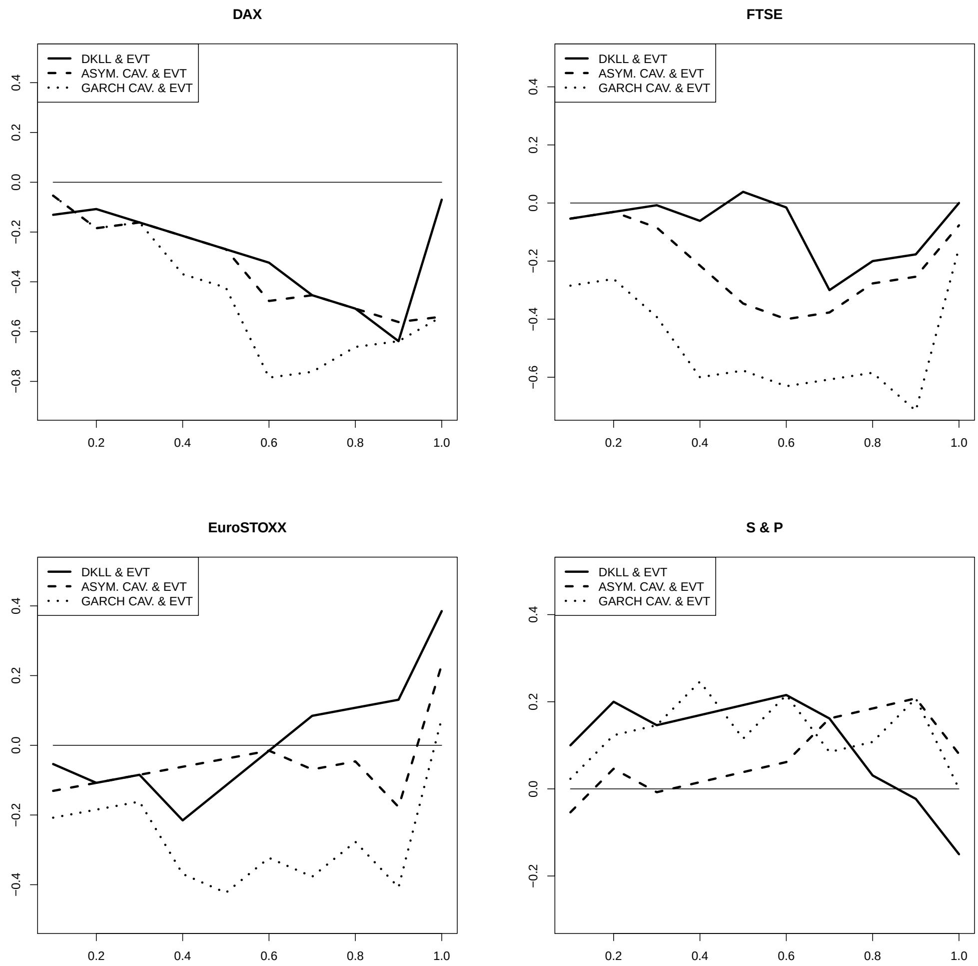

Figure 5.4: Coverage results in % for 0.1%-1% ranges of estimated EVT-augmented index VaRs. The evaluated forecast horizon is 1300 days. Nominal coverages are on the horizontal axis, and the lines correspond to the differences of nominal and estimated VaR exceedance shares (in %). The closer they are to zero, the better the (unconditional) model fit.

We adapt their graphical representation of coverage accuracy in Figure 5.4 for VaR levels between 0.1% and 1%. For the sake of clarity, the AR-TGARCH CAViaR, which usually showed results that were similar to one of the other two CAViaR models, is not included in the graph. It turns out that the Asymmetric Slope CAViaR is the stronger competitor for the DKLL model, as in three out of four cases, the lines corresponding to its differences to the nominal coverages are closer to zero than those corresponding to the GARCH CAViaR. In some ranges, e.g. 0.1%-0.5% FTSE VaR, the DKLL model clearly yields a very good fit. In other cases, such as DAX, all three models do not hit the correct coverages. The DKLL estimator, however, goes head to head with the CAViaR models, while sometimes even beating them.

6 Simulation: Comparing DKLL and EVT-refined VaR

This section is devoted to the question of whether it is sensible to refine the nonparametric VaR estimator with extreme value methods, instead of using the plain version even for extreme VaR estimation. One would expect that especially for small data sets, the EVT extrapolation into the far tails of return distributions yields more stable results than estimating the tail quantiles directly. In order to check this, we carry out a small simulation study to complement the empirical results from the previous sections. As the goal is to assess the relative accuracies of DKLL and EVT-refined DKLL estimators, we do not additionally include the CAViaR models.

To the FTSE time series, we fit a GARCH(1,1) model with t-distributed error terms. It has the following form:

$$Y_t = \mu + \sigma_t \epsilon_t, \;\; \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \;\; \epsilon_t \sim t_\nu. \tag{6.1}$$

Parameter estimates are listed in Table 6.1.

$\widehat{\mu}$	$\widehat{\omega}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\nu}$
0.054	0.015	0.083	0.904	10

Table 6.1: Estimated GARCH parameters from the FTSE return series with 4998 observations. This estimated model is used in the simulation.

From the estimated model, a time series of 13000 observations is simulated. To obtain a setup which is realistic with respect to usual data availability, only 2000 observations are used to estimate the models, and 0.1% VaR predictions are computed over two forecast horizons, N=5000 and N=10000. The advantages of simulated data are that they allow for much longer horizons, and that the return quantile functions

$$q_p^t = \sigma_t F_p^{-1}(\epsilon_t),$$

where $F_p^{-1}(\epsilon_t)$ denotes the p-quantile of the error term distribution, can be computed because the input parameters are known. This allows us to compare the estimated VaRs to their true counterparts. Table 6.2 shows coverages, mean squared errors, mean absolute errors and median absolute errors in-sample and out-of-sample for both models.

	in-sample: n=2000			
	cov.	$\widehat{MSE}$	$\widehat{MAE}$	$\widehat{Med.AE}$
DKLL	0.001	1.049	0.753	0.52
EVT-DKLL	0.001	0.605	0.579	0.429

	out-of-sample: N=5000				out-of-sample: N=10000			
	cov.	$\widehat{MSE}$	$\widehat{MAE}$	$\widehat{Med.AE}$	cov.	$\widehat{MSE}$	$\widehat{MAE}$	$\widehat{Med.AE}$
DKLL	0.009	3.673	1.359	0.9	0.008	3.379	1.253	0.787
EVT-DKLL	0.003	2.455	1.067	0.649	0.004	2.281	0.981	0.578

Table 6.2: Coverages and different loss functions from comparing the estimated 0.1% VaRs with the true quantile function. Cov. stands for coverage, MSE for mean squared error, MAE for mean absolute error and Med. AE for median absolute error.

Throughout, the EVT-augmented DKLL model yields lower losses and better coverages than the plain DKLL model. In order to robustify this result, we repeated the simulation for GARCH parameters estimated from EuroSTOXX data, and using different numbers of in-sample observations. All these results, which are available on request, lead to the conclusion that that the combination of standardized nonparametric residuals and extreme value theory is a valuable complement to the rearranged DKLL estimator, which we suggest to use for estimating moderately low quantiles.

7 Conclusion

In this paper, we propose a way to estimate and predict conditional Value at Risk from a nonparametric model. We consider probabilities that are of practical interest for financial institutions. For external market risk reporting, 1% portfolio VaRs have to be estimated on a daily basis. Internal risk management sometimes requires to take into account even more extreme probabilities such as 0.1%. Although typically very few observations are available in the extreme tails, models to be used should be flexible and rest upon as few structural assumptions as possible. We suggest to use nonparametric quantile regression, more specifically, a rearranged Double Kernel Local Linear VaR estimator as well as a version of the latter augmented by extreme value theory. Both are applied to different index return time series. Forecast performances are benchmarked against the widely used CAViaR models. Although these also perform well in many occations, none of the considered specifications systematically dominates the others. In constrast to them, nonparametric regression circumvents the issue of choosing the appropriate parametrization. Backtesting results from the evaluation of real as well as simulated data examples lead to the conclusion that the fully nonparametric and the EVT-refined nonparametric models do not only outperform the parametric alternatives in a considerable number of situations, but that they can be used to predict VaR of any probability level of interest, even when the estimation period is of moderate size. In recent years, computing power has increased to such an extent that fully nonparametric models come at little more computation cost than other models that rely on more restrictive assumptions. From the results in this paper, however, we conclude that the gains on the additional flexibility are substantial and nonparametric quantile regression with EVT refinements should be considered as a practical alternative for estimating and forecasting VaR.

Acknowledgements

I would like to thank Melanie Schienle and Nikolaus Hautsch 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.

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.

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 (2009). Improving point and interval estimates of monotone functions by rearrangement. *Biometrika 96*, 559–575.

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

Dette, H., N. Neumeyer, and K. F. Pilz (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli 12*, 469–490.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli 9*, 617–657.

El-Arouia, M.-A. and J. Diebolt (2002). On the use of the peaks over thresholds method for estimating out-of-sample quantiles. *Computational Statistics and Data Analysis 39*, 453–475.

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.

Escanciano, J. C. and J. Olmo (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business and Economic Statistics 28*, 36–51.

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.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 73–84.

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

MacDonald, A., C. Scarrott, D. Lee, B. Darlowb, M. Reale, and G. Russell (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis 55*, 2137–2157.

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.

McNeil, A. J., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

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.