Nonparametric Quantile Regression in Time Series

Julia Schaumburg*

April 1, 2009

*juliaschaumburg@web.de

1 Introduction

(Conditional) Value at Risk is a standard measure of market risk, used by financial institutions as well as external regulators such as the Basle Committee of Banking Supervision. It corresponds to a particular quantile of future loss distribution: Given the available information, for each period a loss is forecast that the current portfolio will exceed with some (low) probability p.

Although the concept is simple, some challenges arise when estimating conditional Value at Risk (VaR): It might not be clear in advance what to assume on the functional form of the relationship between VaR and regressors. Furthermore, since the true value of a quantile is not observed and can therefore not be used as benchmark, one has to come up with a method of forecast evaluation. And thirdly, the information set conditioned on has to be chosen in a meaningful way.

This paper aims at discussing some aspects of the first of these issues: In order to gain insight on possible functional forms of VaR curves, fully nonparametric regression is carried out, requiring only weak conditions on the true functions and return distributions. In advance, two candidate nonparametric estimators, the Double Kernel Local Linear (DKLL) quantile estimator of Yu and Jones (1998) and the Weighted Nadaraya Watson (WNW) estimator of Hall et al. (1999) and Cai (2002) are compared in a small simulation study, on the basis of which the DKLL estimator is chosen for the VaR application. For three time series of index returns, nonparametrically estimated quantile curves as functions of lagged returns are visually compared to news impact curves obtained from a new variant of the nonlinear autoregressive Conditional Autoregressive Value at Risk (CAViaR) models of Engle and Manganelli (2004), which allows for asymmetric nonlinear shapes of conditional VaR curves.

Although in-sample fits are an informative first step, in practice the task of VaR forecasting, i.e. the assessment of potential future losses, is of particular interest. It turns out that the DKLL estimator, which smoothes not only the regressor observations, but also the dependent observations, leads to 5 % VaR curves almost without distortions in the sparse regions of the support, except for the outmost boundaries. Therefore, forecasts can be computed and tested for accuracy by the dynamic quantile out-of-sample forecast test of Engle and Manganelli (2004). The forecasts are benchmarked by two existing and the new CAViaR specifications.

Section 2.1 introduces the fully nonparametric WNW and DKLL estimators, before comparing them with respect to their ability to fit conditional quantiles to simulated data. Based on the results, the DKLL estimator is chosen as VaR estimator. Section 2.2 outlines the modelling idea of the nonlinear parametric CAViaR models of Engle and Manganelli (2004). Furthermore, the new Indirect T(hreshold) GARCH(1,1) CAViaR is introduced, which allows for nonlinearity as well as asymmetry of estimated quantile curves. VaR estimation results are described in section 3.1, while DKLL estimator and three CAViaR specifications are used for VaR forecasting in section 3.2.

2 Quantile Models

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.[1] The 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, for versions of (2.2) see for example the seminal paper by Koenker and Bassett (1978); furthermore also Cai and Xu (2008), Engle and Manganelli (2004), Chernozhukov (2005), Yu and Jones (1997) or Koenker and Zhao (1996). On the other hand, Cai (2002), Yu and Jones (1998), Wu et al. (2007) or Chernozhukov and Umantsev (2001) operationalize (2.1).

Although the two formulations both define the same unique conditional quantile, estimates based on (2.1), on the contrary to the ones derived from sample equivalents of (2.2), are prevented from crossing each other, a feature arising from monotonocity of conditional distribution estimates, as pointed out by Cai (2002) and Koenker (2005).

Following the convention of expressing it as a positive number, VaR, corresponding to probability p, at time t is defined as

$$VaR_p^t = -q_p(\mathbf{x}),$$

where VaR_p^t denotes a generic VaR measure.[2]

[1] For the respective model specific assumptions, see the original papers by Cai (2002), Yu and Jones (1998) and Engle and Manganelli (2004).

[2] For the VaR estimators defined below, different notation is used in order to distinguish them from each other.

2.1 Nonparametric Models

The derivations of Weighted Nadaraya Watson (WNW) and Local Linear Double Kernel (DKLL) estimator are based on sets of bivariate observations $\{(X_t, Y_t)\}_{t=1}^n$ drawn from some underlying distribution $F(x, y)$ with density $f(x, y)$. The estimators are set up as inverse of conditional distribution functions: A collection of quantiles at given covariate value $X_t = x$ is derived.[3]
A nonparametric method of estimating a conditional distribution $F(y|x)$ is

$$\check{F}(y|x) = \sum_{t=1}^{n} w_t(x) I(Y_t \leq y), \tag{2.3}$$

where $w_t(x)$ are greater than zero and sum up to one. Choosing equal weights $w = 1/n$ delivers the empirical distribution function. Using instead a kernel function

$$K\left(\frac{X_t - x}{h}\right),$$

in the following abbreviated by $K_h(\cdot) = 1/hK(\cdot/h)$, which is often chosen to be a bounded symmetric density, to attach a smooth set of weights

$$w_t(x) = \frac{K_h\left(X_t - x\right)}{\sum_{t=1}^{n} K_h\left(X_t - x\right)}$$

to the data, results in the Nadaraya Watson (NW) estimator for conditional distribution (see for example Li and Racine (2007), p. 182). It is known to be monotone increasing and bounded between zero and one, but it suffers from boundary distortion, as pointed out by Fan and Gijbels (1996). They advocate the use of local polynomial estimators, the simplest of which is the local linear (LL) estimator. These, however, do not possess the advantage of producing monotone increasing, bounded conditional distribution estimates, which may cause difficulties when inverting them to obtain conditional quantiles, the sample equivalents of (2.1). The two estimators presented below adress these issues in two different ways. In order to choose the one which is more appropriate for VaR forecasting, a small simulation study is carried out.

Weighted local constant smoother The idea of the WNW estimator is to choose a set of weight functions $p_t(x)$ such that they fulfill the discrete moment conditions of the simplest local polynomial estimator, the local linear. Fan and Gijbels

[3]Throughout this section, quantiles, not VaR are discussed. For the empirical analysis in section 3, VaR corresponds to the negative quantile.

(1996)[4] showed that these moment conditions ensure design adaptivity of local polynomial estimators. However, they do not hold for the 'ordinary' NW estimator. The two conditions are

$$\sum_{t=1}^{n} p_t(x) = 1 \quad \text{and} \quad \sum_{t=1}^{n} p_t(x)(X_t - x)K_h(X_t - x) = 0. \tag{2.4}$$

Functions $p_t(x)$ fulfilling these conditions are not unique. One possibility to identify them distinctly 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 (2.4). After taking derivatives and rearranging, $p_t(x)$ simplify to

$$p_t(x) = n^{-1}\left[1 + \lambda(X_t - x)K_h(x - X_t)\right] \tag{2.5}$$

Plugging (2.5) into the (Lagrange) objective function gives

$$L = \frac{1}{nh}\sum_{t=1}^{n} log\left[1 + \lambda(X_t - x)K_h(x - X_t\right], \tag{2.6}$$

which is maximized by finding the root of $L'(\lambda) = 0$ numerically, e.g. by Newton's Method. The obtained parameter λ is plugged back into (2.5), which gives the unique weights.

Thus, the WNW estimator for conditional distribution is defined as follows:

$$\hat{F}\left(y|x\right) = \sum_{t=1}^{n} \frac{p_t(x)K_h\left(X_t - x\right)}{\sum_{t=1}^{n} p_t(x)K_h\left(X_t - x\right)} I(Y_t \leq y), \tag{2.7}$$

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

$$\hat{q}_p(x) = \inf\{y \in \Re : \hat{F}(y|x) \geq p\} \equiv \hat{F}^{-1}(p|x) \tag{2.8}$$

It always exists[5] because $\hat{F}(y|x)$ is between zero and one and monotone in y. In Cai (2002) 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. Since the aim of this paper is an assessment of nonparametric Value at Risk estimation and forecasting, the WNW estimator is a feasible candidate, because it is explicitly set up for time series data, and it possesses the mentioned theoretical advantages.

[4] page 63 and 108

[5] See Cai (2002), p. 176.

Double kernel local linear smoother In the original article by Yu and Jones (1998), observations are taken to be independent. However, in simulation it becomes clear that this assumption is not crucial.[6] As the name 'Double Kernel' suggests, not only are the covariate observation weighted by a kernel $K_h(\cdot)$, but smoothing is also carried out in the 'y-direction', in other words, observations of the dependent variable y are localized as well. This requires the introduction of a second symmetric kernel $W_{h_2}(\cdot)$ which has distribution function $\Omega(\cdot)$ and satisfies

$$\int_{-\infty}^{y} W_{h_2}(Y_t - u)du = \Omega\left(\frac{y - Y_t}{h_2}\right) \tag{2.9}$$

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

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0,\beta_1} \sum_{t=1}^{n} \left(\Omega\left(\frac{y - Y_t}{h_2}\right) - \beta_0 - \beta_1(X_t - x)\right)^2 K\left(\frac{x - X_t}{h_1}\right). \tag{2.10}$$

Obviously, the explicit expression for the conditional distribution function estimator,

$$\tilde{F}(y|x) = \sum_{t=1}^{n} \frac{K\left(\frac{x-X_t}{h_1}\right)[S_2 - (x - X_t)S_1]}{\sum_{t=1}^{n} K\left(\frac{x-X_t}{h_1}\right)[S_2 - (x - X_t)S_1]} \Omega\left(\frac{\tilde{y} - Y_t}{h_2}\right), \tag{2.11}$$

where

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

is a form of (2.3) where the distribution function $\Omega(\cdot)$ in (2.9) can be viewed as a smoothed version of the indicator function used for the WNW estimator. It leads to differentiability of conditional distribution function estimates. The indicator function, on the other hand, produces discontinuous estimates. As will be seen in section 3.1, quantile curves estimated by the DKLL are relatively smooth, a feature that is desirable for VaR applications when sparse data regions are present. The double-kernel quantile estimator $\tilde{q}_p(x)$, the sample analogue to (2.1), is defined as follows:

$$\tilde{q}_p(x) = \inf\left\{y \in \Re : \tilde{F}(y|x) \geq p\right\} \equiv \tilde{F}^{-1}(p|x). \tag{2.12}$$

[6]See the following paragraph.

As mentioned at the beginning of this section, $\tilde{F}(y|x)$ might not always be monotone increasing. In this case, the implementation scheme of Yu and Jones (1998) is adapted: For $\tilde{q}_{1/2}(x)$, any value satisfying (2.12) is chosen; for $p > 1/2$, the largest, and for $p < 1/2$, the smallest solutions to (2.12) are taken as quantile estimates. This procedure ensures monotonicity of the distribution estimate, and therefore circumvents crossing of quantiles. Estimated values > 1 are discarded.

Comparison by simulation An ARCH(1) process with starting value 0 is generated according to

$$Y_t = -0.4X_t + \underbrace{\sqrt{0.4(1 + X_t^2)}}_{\sigma(X_t)}\,\epsilon_t, \tag{2.13}$$

where $X_t = Y_{t-1}$, and the error term $\epsilon_t \sim iidN(0, 1)$. Conditional quantiles are estimated with both WNW ($\hat{q}(x)$) 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 WNW estimator and local linear fit of the DKLL estimator, the gaussian kernel is used, while the uniform kernel is used for smoothing of the dependent variable of the DKLL estimator. Table A.1 contains the results on approximate Integrated Square Error (ISE). For some function $f(x)$, and its estimator $\hat{f}(x)$, ISE is defined as

$$ISE = \int_a^b (\hat{f}(x) - f(x))^2 dx, \tag{2.14}$$

Here, median, 5%, 25%, 75% and 95% quantiles of $I\hat{S}E$ are computed. All results are derived using 500 replications. The tables confirm the consistency results stated in the literature:[7] As sample size increases, all quantiles of ISE as well as the spreads become smaller. It turns out that for the grid point relatively close to the process mean, $x = -0.75$, both estimators give similar results, although the fits derived from DKLL estimates more often give slightly lower ISEs. However, for $x = 1.25$, where data points are sparse while the bandwidth is the same, WNW estimates are notably farther away from the true quantile curves.
Based on these simulation results, the DKLL estimator is chosen as nonparametric VaR estimator, despite the fact that until now it has not been used for time series data, and although it is not automatically monotone and bounded between zero and one.

[7] See the original articles by Yu and Jones (1998) and Cai (2002).

2.2 CAViaR Models

The class of Conditional Autoregressive Value at Risk (CAViaR) models, first introduced by Engle and Manganelli (2004), has since been used as VaR benchmark in a several studies, for example Kuester et al. (2006) or Taylor (2008). A generic CAViaR type model describes 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 lagged values of observable variables, $\mathbf{X}_t$:

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

where $p = q + r + 1$ is the dimension of $\boldsymbol{\beta}$, the parameter that solves

$$\min_{\boldsymbol{\beta}} \frac{1}{T} \sum_{t=1}^{n} \left[p - I\big(Y_t < -VaR_p^t(\boldsymbol{\beta}, \boldsymbol{X}_t)\big)\right](Y_t + VaR_p^t(\boldsymbol{\beta}, \mathbf{X}_t)). \tag{2.15}$$

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. Indeed, the specifications used here include the first lagged value of $VaR_p(\cdot)$ and the first lagged value of Y_t, therefore $\mathbf{X}_t = Y_{t-1}$. Engle and Manganelli (2004) propose four specifications: Adaptive, Symmetric Absolute value, Asymmetric slope and Indirect GARCH(1,1) CAViaR. In the empirical analysis carried out here, only Asymmetric Slope,

$$VaR_p^t(\boldsymbol{\beta}) = \beta_1 + \beta_2 VaR_p^{t-1}(\boldsymbol{\beta}) + \beta_3 (Y_{t-1})^+ + \beta_4 (Y_{t-1})^-, \tag{2.16}$$

where $(x)^+ = \max(x, 0)$ and $(x)^- = -\min(x, 0)$ and indirect GARCH(1,1),

$$VaR_p^t(\boldsymbol{\beta}) = \sqrt{\beta_1 + \beta_2 (VaR_p^{t-1})^2(\boldsymbol{\beta}) + \beta_3 Y_{t-1}^2}, \tag{2.17}$$

are considered. The reason is that the Aymmetric Absolute Value model is just a special case of the Asymmetric Slope model where slope parameters β_3 and β_4 are assumed to be equal. The Adaptive CAViaR specification is motivated such that VaR jumps upwards if a violation occurs, and decreases slightly if no there is no violation. However, the magnitude of Y_{t-1} is not taken into account. This seems to be a too strong simplification, and therefore, only the more flexible above specifications are considered.

When analyzing volatility, two stylized facts are often detected: Firstly, a high positive or negative return on one day often entails another high return for the next day, i.e. there is volatility clustering. This 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 with respect to the sign of past returns. This possibility is captured by the Asymmetric Slope specification (2.16), but not by the Indirect GARCH(1,1) specification. On the other hand, the Asymmetric Slope CAViaR imposes a piecewise linear structure on VaR, although the true functional form might be nonlinear. Thus, a straightforward extension to the existing models is an Indirect TGARCH(1,1) specification, where T stands for threshold:

$$VaR_p^t(\boldsymbol{\beta}) = \sqrt{\beta_1 + \beta_2(VaR_p^{t-1})^2(\boldsymbol{\beta}) + \beta_3 Y_{t-1}^2 + \beta_4(Y_{t-1})^2 I(Y_{t-1} < 0)}, \qquad (2.18)$$

It allows to account for nonlinearity as well as for asymmetry.

3 Empirical Analysis

3.1 VaR Estimation

Cai and Wang (2008) introduce nonparametric VaR and expected shortfall estimators and compare them visually to CAViaR news impact curves for a sample of S&P 500 index returns. However, they plot only a small range, namely $[-1\%, 1\%]$. Stock index returns typically have larger variation, although data are sparse at the boundary, which may distort nonparametric estimates. For forecasting, such distorted areas are useless. Due to the additional smoothing, however, it turns out that the DKLL quantile estimator to a certain extend is able to deal with data sparseness. Therefore, it can be used for VaR forecasting (see section 3.2).

The data set, covering February 2, 1999 to April 25, 2008 consists of 2410 observations of 100 times daily log returns of Euro STOXX, S&P 500 and FTSE 500. In-sample fits are computed using the first 2110 observations, while 300 observations are left for out-of-sample forecasting. 5% VaR curves are estimated by the DKLL estimator as well as Asymmetric slope, Indirect GARCH(1,1) and Indirect TGARCH(1,1) CAViaR estimators.

Figures A.1, A.2 and A.3 show graphs of DKLL and TGARCH(1,1) news impact curves, i.e. the 'reactions' of VaR to different magnitudes of lagged returns. The DKLL estimator does not impose any specific functional form, so that it can be viewed as an indication to the actual features of the data sets. For the CAViaR model, $VaR_p^{t-1}(\boldsymbol{\beta})$ is fixed at its mean.

A few attributes are shared by all news impact curves, for all the data sets: Firstly,

there is strong evidence for nonlinearity in coefficients. In particular, volatility clustering seems to be present, suggesting that the effects of lagged returns on VaR increase, the further away returns are from their empirical average. Secondly, these effects are not symmetric: Large positive returns do not increase risk to the same extend as large negative returns. Thus, the inclusion of the asymmetry parameter in the Indirect TGARCH(1,1) specification seems to reveal some additional information compared to the symmetric Indirect GARCH(1,1) model. The question of whether the TGARCH model describes the data better than the Asymmetric Slope model, cannot be easily answered by looking at the graphs, however. It is left to the comparison of forecast performance.
For EuroSTOXX returns, DKLL and TGARCH news impact curves basically exhibit the same shape, with the difference that the former lies entirely above the latter. The percentage of in-sample VaR exceedances (coverage) of the DKLL estimator for EuroSTOXX, listed in table A.2, confirms the finding that the nonparametric estimator slightly overestimates VaR: The in-sample coverage amounts to 4.77%, whereas it is above 5% for the Indirect TGARCH CAViaR.
The two FTSE news impact curves almost lie on top of each other on the interval $[-2\%, 0.1\%]$, which is an indication that for negative lagged returns, the TGARCH CAViaR describes the data quite well. For lagged returns above zero, on the other hand, the DKLL estimate indicates that there are upward as well as downward movements, which are averaged by the TGARCH model, resulting in a slope that is close to zero. This does not necessarily imply that forecast performance of the DKLL estimator is better, however; too sensitive in-sample fits may lead to poor forecast performance.
Figure A.3, which shows news impact curves for the S&P data, suggests that both models have limits, for different reasons: Being centered at zero, the TGARCH model averages a down- and then upward movement, such that a negative coefficient results. The DKLL estimate indicates, however, that risk increases for large positive returns, but that the minimum is above zero, which is likely to be the reason for the negative TGARCH coefficient. On the other hand, the DKLL curve is 'edgier' and seems to exhibit more distortions than the estimates on the other figures; this may result in less precise VaR forecasts.
Estimated DKLL curves are available almost for the entire support of the return data, except the outmost boundaries.[8] The example of S&P 500 however, shows that despite the 'double smoothing', distortions may be present. A possible way

[8]WNW estimates, on the other hand are severely distorted even in moderately sparse areas, so that they can hardly be used for forecasting. The corresponding figures are not reported here.

to attack this difficulty consists in the choice of an entire set of different bandwidths rather than only two. This task is left for future research.

3.2 VaR Forecasting

The realized quantiles of process Y_t are not observed, and cannot be compared to the forecasted values. Therefore, following Engle and Manganelli (2004), backtesting of VaR models is carried out by defining a function

$$Hit_t = I(Y_t < -VaR_p^t) - p \tag{3.1}$$

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

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

Summarizing the above points, VaR is estimated correctly, if for each day independly, the probability of exceeding it equals p. For setting up the test, a regression equation

$$Hit_t = \mathbf{X}_t'\boldsymbol{\theta} + u_t, \quad u_t = \begin{cases} -p & \text{with prob. } 1-p \\ 1-p & \text{with prob. } p \end{cases} \tag{3.2}$$

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

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

is tested by a Wald test for joint significance, which has test statistic

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

where n_{OOS} is the number of out-of sample forecasts, n_{IS} is the number of observations used for estimating the model, and $\mathbf{Hit}$ and $\mathbf{X}$ are the vector with observations of the dependent variable and the regressor matrix, respectively.

Following common practice in the literature,[9] the information set here consists of a constant, four lagged values of Hit_t and the respective estimate of VaR_p^{t-1}. Results on in-sample and out-of-sample coverage as well as p-values from the DQ test are summarized in table A.2. It turns out that the test does not detect severe misspecification for any of the models: All p-values are 0.1 or above, so it can be concluded that there are no periods where VaR violations cluster. In-sample coverage in all cases is close to 5%, which is not surprising for the CAViaR models, because the objective function (2.15) ensures that the parameters are chosen in exactly this way. However, the DKLL estimator achieves a good in-sample fit as well.

Results on out-of-sample forecasting performance, on the other hand, do not lead to a clear indication of which model performs best. Coverages are close to 5% for all models when applied to the FTSE and the EuroSTOXX data. For the S&P 500 returns, VaRs are of the forecasting period are underestimated by all four models. The asymmetric CAViaR models exhibit negative coefficients for positive returns.[10] The reason might be that for both Asymmetric Slope and Indirect TGARCH(1,1), the threshold is zero, i.e. the value for which a change of coefficients is assumed, is fixed. From visual inspection of the DKLL curves, this assumption works out for the EuroSTOXX and the FTSE data, but in case of S&P, averaging all returns greater than zero for estimation of the 'right hand-side coefficient' apparently results in a misspecification bias, which is confirmed by the poor forecast performance.

The DKLL forecasts themselves are not considerably more accurate than the CAViaR forecasts. This might be due to the above mentioned remaining distortions, that are present despite the double-smoothing. On the other hand, the DKLL estimator relies on the assumption that once they are estimated, return distributions do not change over time, whereas CAViaR is a distribution-free approach, and it is therefore more flexible in this respect.

4 Conclusion and Outlook

Fully nonparametric quantile regression, although computationally involved, allows insights on the relationship between regressors and Value at Risk without placing specific assumptions on functional form in advance. For the investigated data sets, visual inspection of the nonparametrically estimated VaR curves sug-

[9] See Engle and Manganelli (2004), Kuester et al. (2006) and Taylor (2008).

[10] The table of CAViaR coefficients is not reported due to space limitations, but it can be seen in Figure A.3 that for TGARCH estimates, VaR goes down as the lagged return increases.

gested that asymmetric volatility clustering was present, justifying the introduction of a new variant of the popular CAViaR models that accounts for nonlinearity as well as asymmetry. This new Indirect TGARCH(1,1) specification showed good forecast performance except for the S&P 500 return series, where it underestimated risk along with the other models. This result points to the conclusion that a threshold of zero, indicating a saddle point or minimum, is inappropriate for some data sets. Therefore, another possible extension of the model would be to include an additional nonzero, time-varying mean term.[11] Then one would not expect negative coefficients anymore, and CAViaR news impact curves might be even closer to their nonparametrically estimated counterparts, which may lead to a decrease of VaR exceedances. For FTSE 100 and Euro STOXX 50, however, where the minimum VaR occured at a return value closer to zero, all considered models performed reasonably well.

In this study, emphasis was placed on finding evidence of how an appropriate shape of VaR curve looks like. The DKLL estimator itself, although it possesses the advantage of additional smoothness compared to other nonparametric quantile estimator such as the WNW estimator, did not lead to a better forecast performance than the CAViaR models. This finding may partly be attributed to remaining boundary distortions due to data sparseness. Two possibilities of improving on this are on the agenda for future research projects: Firstly, one could come up with a set of bandwidths that adapts to data sparseness. And secondly, the kernel distribution used for smoothing of the dependent observations, could be replaced by a distribution that takes tail estimation into account explicitly.

And finally, the DKLL estimator has been used here as if it had been set up for time series data, which originally was not the case. The theoretical result that it does indeed work under dependence, is about to appear.

[11] See also Kuester et al. (2006).

A Appendix

$x = -0.75$

	5% ISE	25% ISE	Median ISE	75% ISE	95% ISE
		n=200			
$\hat{q}_p(x)$	0.0247	0.0351	0.0483	0.0700	0.1142
$\tilde{q}_p(x)$	0.0276	0.0345	0.0434	0.0559	0.0817
		n=500			
$\hat{q}_p(x)$	0.0172	0.0215	0.0279	0.0379	0.0626
$\tilde{q}_p(x)$	0.0180	0.0232	0.0282	0.0341	0.0473
		n=1000			
$\hat{q}_p(x)$	0.0116	0.0159	0.0194	0.0254	0.0382
$\tilde{q}_p(x)$	0.0148	0.0181	0.0213	0.0248	0.0321

$x = 1.25$

	5% ISE	25% ISE	Median ISE	75% ISE	95% ISE
		n=200			
$\hat{q}_p(x)$	0.0722	0.1152	0.1723	0.2722	0.5620
$\tilde{q}_p(x)$	0.0595	0.0808	0.1044	0.1363	0.2463
		n=500			
$\hat{q}_p(x)$	0.0412	0.0617	0.0847	0.1220	0.2576
$\tilde{q}_p(x)$	0.0362	0.0515	0.0628	0.0802	0.1180
		n=1000			
$\hat{q}_p(x)$	0.0326	0.0446	0.0594	0.0773	0.1169
$\tilde{q}_p(x)$	0.0287	0.0374	0.0442	0.0532	0.0713

Table A.1: ISEs for different sample sizes, conditioning on $x = -0.75$ and $x = 1.25$. WNW Quantile Estimator $\hat{q}_p(x)$ and DKLL Quantile Estimator $\tilde{q}_p(x)$.

EuroSTOXX 50				
	Asymmetric Slope	GARCH	TGARCH	DKLL
In-sample coverage	0.0524	0.0524	0.0529	0.0477
Out-of sample coverage	0.0526	0.0676	0.0563	0.0381
DQ Test: p-value	0.74	0.54	0.19	0.13

S&P 500				
	Asymmetric Slope	GARCH	TGARCH	DKLL
In-sample coverage	0.0534	0.0534	0.0534	0.0498
Out-of sample coverage	0.0989	0.0791	0.0830	0.0869
DQ Test: p-value	0.12	0.46	0.20	0.23

FTSE 100				
	Asymmetric Slope	GARCH	TGARCH	DKLL
In-sample coverage	0.0529	0.0534	0.0529	0.0482
Out-of sample coverage	0.0563	0.0563	0.0563	0.071
DQ Test: p-value	0.32	0.11	0.17	0.10

Table A.2: DQ test results as well as in-sample and out-of sample coverage of VaR forecasts (percentage of out-of-sample hits).

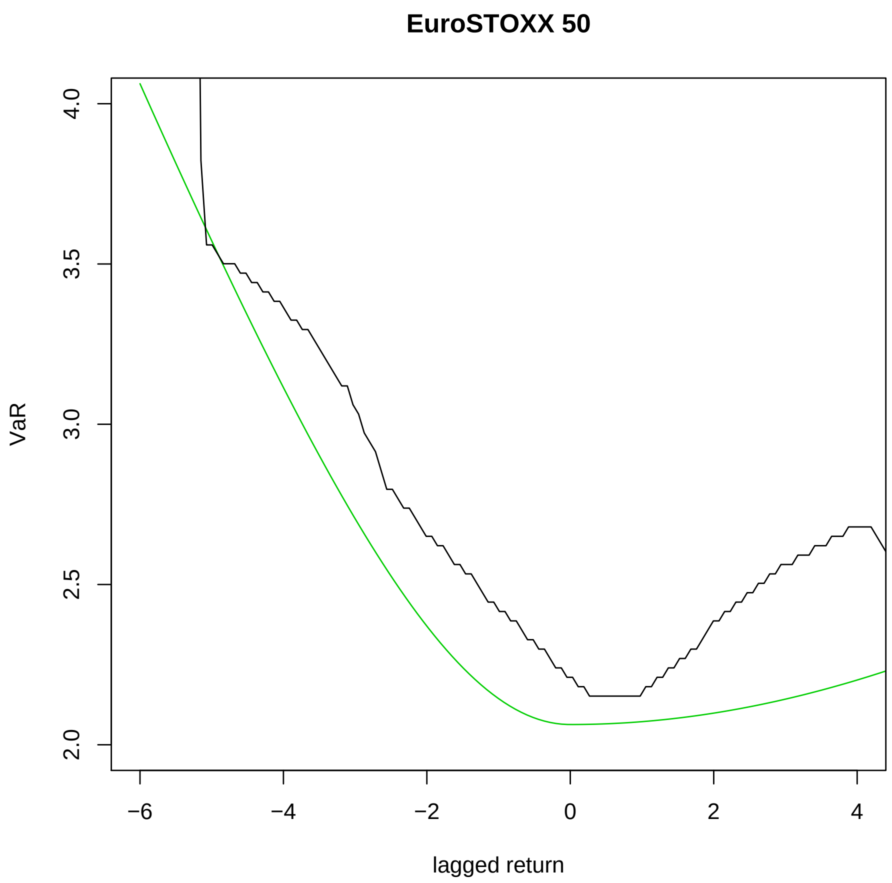

Figure A.1: EuroSTOXX 50

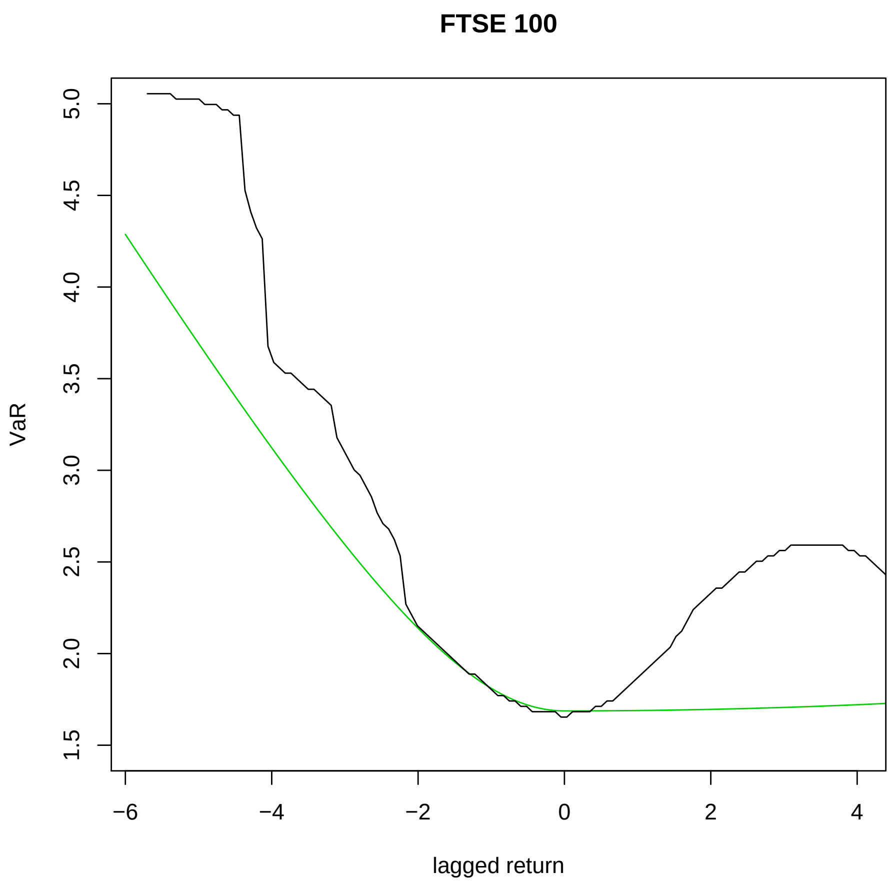

Figure A.2: FTSE 100

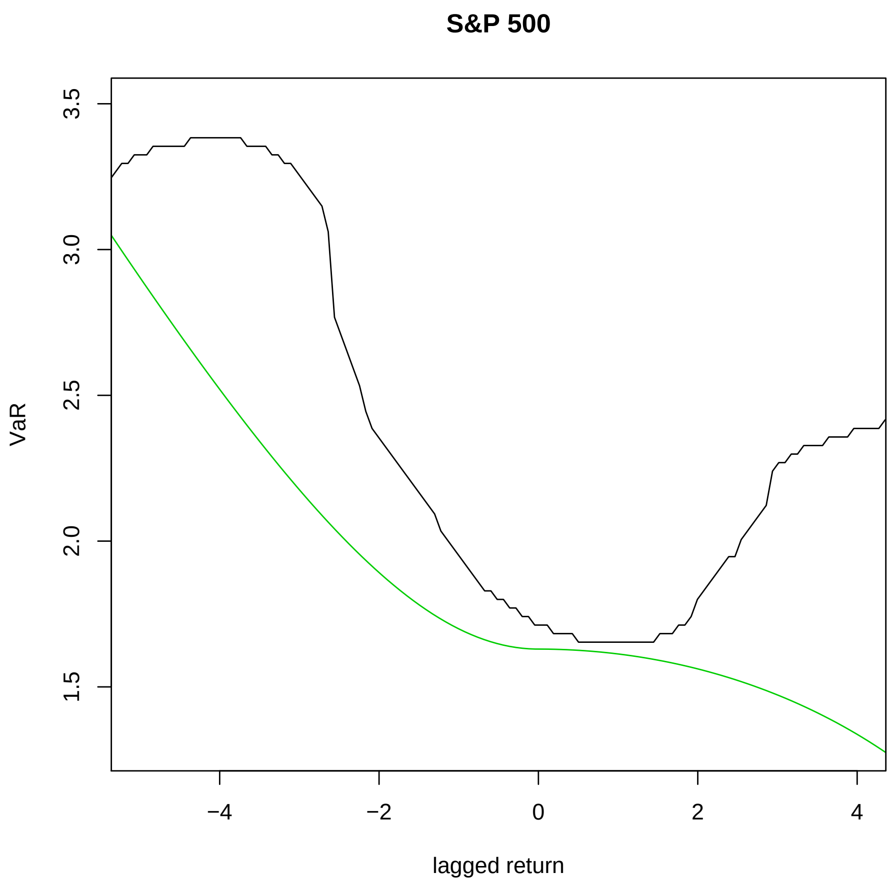

Figure A.3: S&P 500

References

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.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*, 806–839.

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

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.

Hall, P., R. C. L. Wolff, and Q. Yao (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association 94*, 154–163.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

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. Paolelle (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.

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.