Expectile and quantile regression—David and Goliath?

Linda Schulze Waltrup[1], Fabian Sobotka[2], Thomas Kneib[2] and Göran Kauermann[1]
[1]Ludwig-Maximilians-Universität Munich, Germany
[2]Georg-August-Universität Göttingen, Germany

Abstract: Recent interest in modern regression modelling has focused on extending available (mean) regression models by describing more general properties of the response distribution. An alternative approach is quantile regression where regression effects on the conditional quantile function of the response are assumed. While quantile regression can be seen as a generalization of median regression, expectiles as alternative are a generalized form of mean regression.

Generally, quantiles provide a natural interpretation even beyond the 0.5 quantile, the median. A comparable simple interpretation is not available for expectiles beyond the 0.5 expectile, the mean. Nonetheless, expectiles have some interesting properties, some of which are discussed in this article. We contrast the two approaches and show how to get quantiles from a fine grid of expectiles. We compare such quantiles from expectiles with direct quantile estimates regarding efficiency. We also look at regression problems where both quantile and expectile curves have the undesirable property that neighbouring curves may cross each other. We propose a modified method to estimate non-crossing expectile curves based on splines. In an application, we look at the expected shortfall, a risk measure used in finance, which requires both expectiles and quantiles for estimation and which can be calculated easily with the proposed methods in the article.

Key words: Expected shortfall, least asymmetrically weighted squares, non-crossing, penalized splines, semiparametric

1 Introduction

Quantile regression allows the estimation of the effect of covariates on the distribution of a response variable. The idea has been suggested by Koenker and Bassett (1978) and is well elaborated with numerous extensions in Koenker (2005). The underlying regression model for the α-quantile with $\alpha \in (0, 1)$ is specified as

$$y_i = q_{i,\alpha} + \epsilon_{i,\alpha}, \quad i = 1, \dots, n$$

with y_i as response variable, $i = 1, \ldots, n$ and $q_{i,\alpha}$ as the α-quantile which may depend on covariates x_i, say, e.g., through the linear model $q_{i,\alpha} = \beta_{0\alpha}^{(q)} + x_i\beta_{1\alpha}^{(q)}$. Unlike classical regression where a zero mean is assumed for the residuals, in quantile regression one postulates that the α-quantile of the residuals $\epsilon_{i,\alpha}$ is zero, i.e., $P(\epsilon_{i,\alpha} \leq 0) = \alpha$. Estimates for $q_{i,\alpha}$ are obtainable through the minimizer of the weighted L_1 sum

$$\sum_{i=1}^{n} w_{i,\alpha} \left| y_i - q_{i,\alpha} \right|, \qquad (1.1)$$

where

$$w_{i,\alpha} = \begin{cases} 1 - \alpha, & \text{for } y_i < q_{i,\alpha}, \\ \alpha, & \text{for } y_i \geq q_{i,\alpha} \end{cases}$$

are asymmetric weights. Numerically, (1.1) can be minimized by linear programming, see, e.g., Koenker (2005).

As an alternative to quantile regression, Aigner *et al.* (1976) and Newey and Powell (1987) proposed to replace the L_1 distance in (1.1) by a quadratic L_2 term leading to the asymmetric least squares

$$\sum_{i=1}^{n} w_{i,\alpha} \left(y_i - m_{i,\alpha} \right)^2 \qquad (1.2)$$

where the minimizer $\hat{m}_{i,\alpha}$, say, is called (estimated) expectile. The underlying regression model now is

$$y_i = m_{i,\alpha} + \epsilon_{i,\alpha}, \quad i = 1, \ldots, n$$

under the assumption that the α-expectile $m_{i,\alpha}$ of the error terms is zero. The expectile $m_{i,\alpha}$ may again depend on covariates, e.g., through the linear expectile model $m_{i,\alpha} = \beta_{0\alpha}^{(m)} + x_i\beta_{1\alpha}^{(m)}$. Expectile estimation is thereby a special form of M-quantile estimation (Breckling and Chambers, 1988) and expectile regression has seen some increasing interest in the last years (Schnabel and Eilers, 2009b; Pratesi *et al.,* 2009; Sobotka and Kneib, 2012; Guo and Härdle, 2013). An overview about methods focusing on estimation procedures regarding more features of the data than its centre (including semiparametric expectile and quantile regression) can be found in Kneib (2013).

In this article, we contrast quantile and expectile regression and propose some extensions to expectile estimation to link it to quantiles. A comparison of the two routines might remind us of the story of David and Goliath just by comparing the number of citations: about 1850 for Koenker and Bassett (1978) referring to quantiles and about 100 for Newey and Powell (1987), as of November 2013. Quantiles are certainly more dominant in the literature due to the fact that expectiles lack an intuitive interpretation while quantiles are just the inverse of the distribution function. Numerically, as seen by comparing (1.1) to (1.2), quantiles 'live' in the L_1 world

while expectiles are rooted in the L_2 world. This by itself has several consequences. Quantiles need linear programming for estimation while expectiles are fitted using quadratic optimization. Beyond all discrepancies between quantiles and expectiles, it is important to note that both are related in various ways. Jones (1994) shows that expectiles are in fact quantiles of a distribution function uniquely related to the distribution of y. Yao and Tong (1996) give a similar result by showing that there exists a unique bijective function $h : (0, 1) \rightarrow (0, 1)$ such that $q_\alpha = m_{h(\alpha)}$, where $h(.)$ is defined through

$$h(\alpha) = \frac{-\alpha q_\alpha + G(q_\alpha)}{-m_{0.5} + 2G(q_\alpha) + (1 - 2\alpha)q_\alpha} \tag{1.3}$$

with $G(q) = \int_{-\infty}^{q} y \, dF(y)$ as the partial moment function and $F(y)$ as cumulative distribution of y (see also De Rossi and Harvey, 2009). Note that $m_{0.5} = E(y) = G(\infty)$. In this article, we will make use of relation (1.3) and relate quantile estimates $\hat{q}_\alpha$ to expectile-based quantile estimates $\hat{m}_{\hat{h}(\alpha)}$, where $\hat{h}(.)$ is an estimated version of $h(.)$ in (1.3). One of the key findings of the article is that $\hat{m}_{\hat{h}(\alpha)}$ are numerically more demanding than quantile estimates, but, as simulations show, they serve as quantile estimates which can be even more efficient than the empirical quantile $\hat{q}_\alpha$ itself.

In quantile regression, a numerical problem in applications are the so-called crossing quantile functions. These occur if for estimated quantiles, one gets $\hat{q}_\alpha(x) > \hat{q}_{\alpha'}(x)$ for $\alpha < \alpha'$ for some value x (in the observed range of the covariate), where $\hat{q}_\alpha(x) = \hat{\beta}_{0\alpha}^{(q)} + x\hat{\beta}_{1\alpha}^{(q)}$. Several methods, algorithms and model constraints have been proposed to circumvent the problem. Bondell *et al.* (2010) made use of linear programming. They also gave a good overview about earlier proposals including Koenker (1984), He (1997), Wu and Liu (2009) or Neocleous and Portnoy (2008). Chernozhukov *et al.* (2010) rearranged the fitted (linear) curves into a set of non-crossing curves, whereas Dette and Scheder (2011) used marginal integration techniques and monotone rearrangements. Schnabel and Eilers (2013b) proposed the so-called quantile sheets where crossings are circumvented by penalization. The problem of crossing curves occurs in principle in the same way for expectile regression. We demonstrate with simulations that crossing expectiles occur less frequently. This implies that less attention is needed to avoid crossing expectiles compared to crossing quantiles.

Quantile regression, as well as expectile regression, can be extended to nonparametric functional estimation. For quantile estimation, Koenker *et al.* (1994) proposed spline-based estimation. Bollaerts *et al.* (2006) made use of penalized B-splines with an L_1 penalty. Recently Reiss and Huang (2012) suggested quantile estimation based on penalized iterative least squares (see also Yuan, 2006). For expectiles, smooth estimation has been pursued by, e.g., Pratesi *et al.* (2009), and Schnabel and Eilers (2009b). Other semi- or non-parametric extensions of quantile regression which allow for flexibility are varying coefficient models. Here smoothness is achieved by letting the quantile not only depend on (unknown) coefficients which are constant, but on (smooth) coefficient functions (see, e.g., Honda, 2004; Noh *et al.,* 2012 or

Andriyana *et al.,* 2014). The idea of smoothing can be extended by assuming that a 'set' of α-quantile curves smoothly depends on both, the covariate and α. Using B-splines this easily allows us to incorporate non-crossing conditions, as in Bondell *et al.* (2010) or Muggeo *et al.* (2013), for quantile estimation. The estimation of quantile sheets is proposed in Schnabel and Eilers (2013b). Schnabel and Eilers (2009a) give a description of non-crossing smooth expectile curves (see also Eilers, 2013). Both approaches are based on penalized spline smoothing. Our approach to estimate expectile sheets is very similar to an extension formulated in Schnabel and Eilers, 2009a, who propose the estimation of expectile sheets using a tensor product of B-splines combined with a penalty in direction of the asymmetry parameter and covariate. But it is different in the sense that we are using a linear B-spline in direction of the asymmetry parameter and avoid the use and choice of a smoothing parameter in that same direction.

Expectiles might not gain popularity as much as quantiles, but we think they deserve their niche. For example, Aigner *et al.* (1976) construct expectiles to estimate production frontiers and give an additional argument for using expectiles by stating that expectile regression is a way to treat asymmetric consequences as it places different weights on positive and negative residuals. But there are other fields which demand for expectiles as well, for example the field of risk measures for financial assets. Ziegel (2013) argues for the use of expectiles as a risk measure as they have desirable properties. Another frequently used risk measure is the 'expected shortfall', which needs the calculation of both, quantiles and expectiles. The expected shortfall (ES) is a trimmed mean, that is, the mean of a random variable conditional that its value is above (or below) a certain quantile. The ES can be written as a function of both, the quantile and the expectile for a level α. Estimation of the ES has been recently proposed by Leorato *et al.* (2012) by employing the integrated (conditional) fitted quantile regression function (see also Wang and Zhou, 2010). We extend an idea of Taylor (2008) and use the fitted quantiles and expectiles for the estimation of the ES. This connection becomes extremely useful for calculating the ES as it depends both on expectiles and their corresponding quantiles (as described by Taylor, 2008).

The article is organized as follows. In Section 2, we compare and contrast quantiles and expectiles, both theoretically and based on simulation. In Section 3, we look at quantile and expectile regression before Section 4 provides extensions and examples. Section 5 concludes the contest of David and Goliath.

2 Expectiles and quantiles

2.1 Quantiles from expectiles

Quantiles as well as expectiles uniquely define a distribution function. Let $F(y)$ denote the continuous distribution function of a univariate random variable Y, which for the sake of simplicity for now is assumed to not depend on any covariates x. The distribution is uniquely defined by the quantile function $q_\alpha = q(\alpha) = F^{-1}(\alpha)$ for $\alpha \in (0, 1)$ or by the expectile function $m_\alpha = m(\alpha)$ for $\alpha \in (0, 1)$. First we show how to

numerically derive the quantile function $q(\alpha)$ from the expectile function $m(\alpha)$. With other words we demonstrate how the transfer function $h(.)$ in (1.3) can be derived numerically, which in practice allows us to calculate the quantile function from a fitted expectile function $\hat{m}(\alpha)$, say. Note first that expectiles are defined through

$$m_\alpha = \frac{(1-\alpha)G(m_\alpha) + \alpha(m_{0.5} - G(m_\alpha))}{(1-\alpha)F(m_\alpha) + \alpha(1 - F(m_\alpha))} \quad (2.1)$$

which needs to be solved numerically for $F(m_\alpha)$. Let, therefore, $0 < \alpha_1 < \alpha_2 \cdots < \alpha_L < 1$ be a dense set of knots covering the $(0, 1)$ interval. In the following we denote $\hat{m}_l = \hat{m}_{\alpha_l}$ to simplify the notation. In principle, a fine grid of expectiles is all we need to estimate the distribution function or quantiles. If the original data is still at hand, one can set $\hat{m}_0 = \min\{y_i, \ i = 1, \ldots, n\} - c_0$ and just as well $\hat{m}_{L+1} = \max\{y_i, \ i = 1, \ldots, n\} + c_{L+1}$, where c_0 and c_{L+1} serve as tuning parameters. In our simulations in Section 2.2 and in the examples in Sections 2.4 and 4.2, we set $\hat{m}_0 = \hat{m}_1 + (\hat{m}_1 - \hat{m}_2)$, $\hat{m}_{L+1} = \hat{m}_L + (\hat{m}_L - \hat{m}_{L-1})$. If one chooses α_1 to be close to zero (and analogously α_L to be close to 1), then there is obviously just a small difference between the minimal expectile and the minimal observed value of the data.

We now solve (2.1) for $\hat{m}_l$, $l = 1, \ldots, L$, and denote the resulting estimator of the cumulative distribution function with $\hat{F}_m(.)$. To obtain $\hat{F}_m(.)$, we estimate the distribution function at the estimated expectiles $\hat{m}_l$ through

$$\hat{F}_l := \hat{F}_m(\hat{m}_l) = \sum_{j=1}^{l} \hat{\gamma}_j \quad (2.2)$$

for non-negative steps $\hat{\gamma}_j \geq 0, j = 1, \ldots, L$ and $\hat{\gamma}_{L+1} = 1 - \sum_{l=1}^{L} \hat{\gamma}_l \geq 0$. Making use of linear interpolation between adjacent values of $\hat{F}(.)$ leads to

$$\hat{G}_l = \hat{G}(\hat{m}_l) = \sum_{j=1}^{l} \hat{c}_j \hat{\gamma}_j$$

with $\hat{c}_j = (\hat{m}_j - \hat{m}_{j-1})/2$ and $\hat{G}_{L+1} = \hat{m}_{0.5}$ as (linear) constraint. This setting now allows to calculate $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_L)$ from a set of estimated expectiles. Details are given in the Appendix.

Defining the linear interpolation $\hat{F}_m(y) = \sum_{j=1}^{l} \hat{\gamma}_j + \hat{\gamma}_{j+1}(y - \hat{m}_l)/(\hat{m}_{l+1} - \hat{m}_l)$ for $y \in [\hat{m}_l, \hat{m}_{l+1})$ allows us to invert $\hat{F}_m(.)$ to obtain quantile estimates based on estimated expectiles. We define these as

$$\hat{m}_{\hat{h}(\alpha)} = \hat{F}_m^{-1}(\alpha).$$

Note that with the definition of $\hat{m}_{\hat{h}(\alpha)}$ for $\alpha \in (0, 1)$, we get an explicit estimate of $\hat{h}(.)$ as a by-product. This is derived by interpolating α_l and $\hat{F}_m(m_l)$, which defines $h^{-1}(.)$, and by taking the inverse we get $\hat{h}(.)$.

We need that $\hat{\gamma}_l \geq 0$ which must be fulfilled since $\hat{m}_l \geq \hat{m}_{l-1}$. Numerical inaccuracy may yield negative values for $\hat{\gamma}_l$, in particular, for α_l close to 0 or 1. Estimation under the linear constraint $\hat{\boldsymbol{\gamma}} \geq 0$ circumvents the problem. Moreover, the estimation can get numerically unstable, which is easily eliminated by imposing a small penalty on the calculated values $\hat{\boldsymbol{\gamma}}$. In fact, defining the density corresponding to $\hat{F}(\cdot)$ as $\hat{f}(\cdot)$ with $\hat{f}(y) = \hat{\gamma}_{l+1}/(\hat{m}_{l+1} - \hat{m}_l)$ for $y \in [\hat{m}_{l+1}, \hat{m}_l)$, we want $\hat{f}(\cdot)$ to be 'smooth'. In other words, $\hat{f}(y) - \hat{f}(y + h)$ should be small for h small. Given that $\hat{f}(\cdot)$ is a step function, this translates to imposing the penalty

$$\lambda_{\text{pen}} \sum_{l=1}^{L-1} \left(\frac{\hat{\gamma}_l}{\hat{m}_l - \hat{m}_{l-1}} - \frac{\hat{\gamma}_{l+1}}{\hat{m}_{l+1} - \hat{m}_l} \right)^2 . \tag{2.3}$$

Details are provided in the Appendix. Note that the calculation of quantiles from expectiles is somewhat numerically demanding. Alternative approaches to estimate quantiles from expectiles are described in Efron (1991) and Schnabel and Eilers (2013a). Schnabel and Eilers (2013a) propose to estimate non-crossing expectile curves using a so-called expectile bundle. Within the expectile bundle crossing curves are prevented but, as a location scale model is assumed, one also loses flexibility. Schnabel and Eilers (2013a) also give a description of how to estimate the density (and therefore, quantiles) from a set of expectiles by using penalized least squares. The approach by Efron (1991) is a more naive way to get an estimation for quantiles on a bases of expectiles. Efron (1991) proposed to estimate a high number of expectiles and to count the number of observations lying below each expectile. He calls the resulting estimates percentiles. Taylor (2008) also uses this method to estimate quantiles from expectiles to calculate the ES. The method proposed in Efron (1991) clearly has the advantage that it is simple and easy to perform, but, as one can imagine, it is not very efficient. Especially for extreme values of α, our method is to be preferred as it leads to more precise estimates.

Conclusion: From a set of expectiles we can numerically obtain the quantile function. The method can also be applied in the regression scenario by conditioning on the explanatory variable as will be demonstrated in the article later.

2.2 Empirical evaluation

Evidently, the resulting fitted distribution function $F_m(.)$ is continuous but has $L + 1$ non-differentiable edges. In principle, one can set L large to n, but this may require heavy and numerically unstable calculations. In our experience, a sequence from $0.0001, 0.001, 0.01, 0.02, \ldots, 0.98, 0.99, 0.999, 0.9999$ usually is sufficient for

deriving the quantile in the range between 1% and 99%, but for large sample sizes, it may be sensible to choose L such that L is proportional to n.

The procedure allows us now to derive quantiles from expectiles and the question arises how they perform in terms of efficiency. We, therefore, run a small simulation study where we estimate a number of expectiles slightly smaller than n (for $n = 499$, we estimated 459 expectiles; for $n = 199$, we set $L = 159$). We simulate (a) from the standard normal distribution, (b) from the Chi-squared distribution (df = 2) and (c) from the t-distribution (df = 3) with sample sizes $n = 199$ and $n = 499$, and each simulation is replicated 1000 times. We use odd sample sizes to guarantee unique quantiles, for e.g., $\alpha = 0.5$. We compare our quantiles from expectiles $\hat{m}_{\hat{h}(\alpha)}$ with ordinary quantiles for $\alpha = 0.01, 0.02, 0.05, \ldots, 0.95, 0.98, 0.99$. The calculation of quantiles from expectiles is a part of the R-package 'expectreg' (as all R-packages available from `cran.r-project.org`). Quantiles $\hat{q}_\alpha$ are calculated using the function `rq` from the R-Package 'quantreg' by Koenker (2013b). We also look at smooth quantiles denoted by $\hat{q}_\alpha^{\text{smooth}}$ and calculated using the method proposed by Jones (1992). For a moderate number of L, numerical instability does not seem to be a problem within the estimation of quantiles from expectiles. In total, the simulation includes 6×1000 times the calculation of quantiles from expectiles, a procedure which failed in none of the 6000 cases. Penalty (2.3) does not only lead to a smooth distribution function and, therefore, to smooth quantile estimates, but also improves the numerical stability of the calculations. For $n = 499$ and a number of 459 expectiles, the implemented function needs around seven seconds to calculate $\hat{F}_m(.)$ (using one kernel of an ordinary computer).

In Figure 1, we show for one (randomly chosen) sample of each distribution with $n = 499$ the fitted transfer function $h(.)$.

The true function is provided for comparison and apparently the fit looks acceptable. The function h itself is of secondary interest for this simulation study, but we see that $h(.)$ in fact can be estimated. Moreover, we will need the transfer function later in the example of Section 4.3 where we estimate the ES.

In Figure 2, we visualize the result of our simulations. Here we compare $\hat{m}_{\hat{h}(\alpha)}$ with $\hat{q}_\alpha$ (solid lines) and, to make a fair comparison, with $\hat{q}_\alpha^{\text{smooth}}$ (dotted lines). As the results for $n = 199$ and $n = 499$ are very similar, Figure 2 concentrates on $n = 199$. The first plot of Figure 2 shows the simulation-based relative root mean squared error (RMSE) of the estimated quantiles for the standard normal distribution. Results for the Chi-squared distribution can be found in the second plot and in the third plot, findings for the student t-distribution are visualized. Results above the unit line stand in favour for expectiles, and Figure 2 mirrors surprisingly satisfactory performances for $\hat{m}_{\hat{h}(\alpha)}$. We notice the gain of efficiency for the two symmetric distributions and inner quantiles: The RMSE for quantiles from expectiles is 5–10% lower than the RMSE for smooth quantiles. Not surprisingly, the difference between quantiles and smooth quantiles becomes stronger when looking at extreme quantiles. This is also mirrored in the relative RMSE as we see that for quantiles reflecting extreme observations the smoothing leads to an improvement. Generally, it occurs that the expectile-based
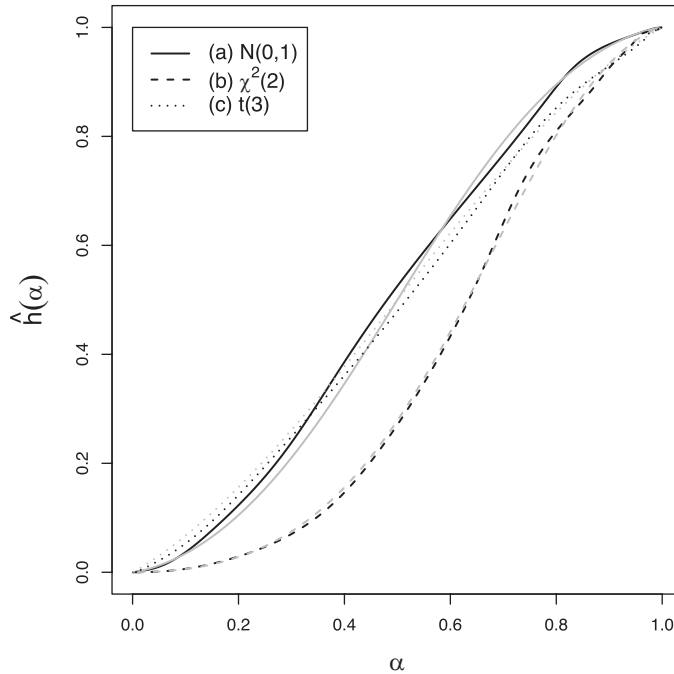
Figure 1 Estimated transfer function $\hat{h}(.)$ (in black) and theoretical transfer function (in grey) for the three kind of error distributions considered here
Source: Authors' own (prepared with GNU R).
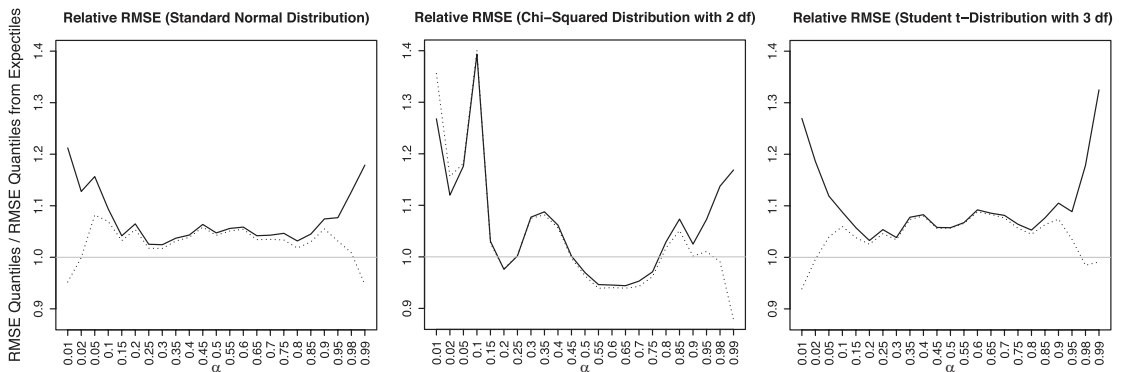
Figure 2 Relative root mean squared error $RMSE(\hat{q}_\alpha)/RMSE(\hat{m}_{\hat{h}(\alpha)})$ (solid lines) and $RMSE(\hat{q}_\alpha^{\text{smooth}})/$ $RMSE(\hat{m}_{\hat{h}(\alpha)})$ (dotted lines) where $h(.)$ is estimated through $\hat{h}(.)$ for different simulation distributions and $n = 199$
Source: Authors' own (prepared with GNU R).

quantile estimators $\hat{m}_{\hat{h}(\alpha)}$ behave sound and are (for most values of α) more efficient than the direct quantile estimates $\hat{q}_\alpha$. This holds as well in terms of relative RMSE as in terms of relative mean absolute error which is not reported here as the results were quite similar.

    <u>Conclusion:</u> All in all, we see, that the calculation of quantiles from a set of expectiles is a sensible thing to do also in terms of efficiency. The numerical burden is of course not ignorable.

2.3 Expectiles and quantiles in the tail

As can be seen from Figure 1, we have for small values of α that $h(\alpha) \ll \alpha$ and accordingly for α close to 1, $(1 - h(\alpha)) \ll (1 - \alpha)$, unless the distribution is heavily tailed. For instance, the $\alpha = 0.01$ quantile of the standard normal distribution corresponds to the $h(\alpha) =: \tilde{\alpha} = 0.0014524$ expectile. This raises the question if and how well extreme expectiles can be estimated. To tackle this question formally, we look at expectiles and quantiles in the tail of the distribution by setting

$$\alpha = \lambda/n \text{ (or } \alpha = 1 - \lambda/n) \tag{2.4}$$

for some $\lambda \geq 1$. Moreover, we assume that the tails have a reasonable interpretation in that the second-order moment of the underlying distribution is finite. The expectile estimate $\hat{m}_{\tilde{\alpha}}$ is defined as a minimizer of (1.2) for $\tilde{\alpha} = h(\alpha)$ and we get

$$\hat{m}_{\tilde{\alpha}} = \left(\sum_{i=1}^{n} \hat{w}_{i,\tilde{\alpha}}\right)^{-1} \left(\sum_{i=1}^{n} \hat{w}_{i,\tilde{\alpha}} Y_i\right), \tag{2.5}$$

where $\hat{w}_{i,\tilde{\alpha}} = 1 - \tilde{\alpha}$ for $Y_i < \hat{m}_{\tilde{\alpha}}$ and $\hat{w}_{i,\tilde{\alpha}} = \tilde{\alpha}$ for $Y_i \geq \hat{m}_{\tilde{\alpha}}$. Note that (2.5) is not an analytic definition, since the iterated weights $\hat{w}_i$ depend on the fitted value $\hat{m}_{\tilde{\alpha}}$. We simplify (2.5) by replacing the 'fitted' weights by their 'true' weights $w_{i,\tilde{\alpha}}$ defined through $w_{i,\tilde{\alpha}} = 1 - \tilde{\alpha}$ for $Y_i < m_{\tilde{\alpha}}$ and $w_{i,\tilde{\alpha}} = \tilde{\alpha}$ otherwise. This allows us to approximate (2.5) to

$$\hat{m}_{\tilde{\alpha}} \approx \left(\sum_{i=1}^{n} w_{i,\tilde{\alpha}}\right)^{-1} \left(\sum_{i=1}^{n} w_{i,\alpha} Y_i\right). \tag{2.6}$$

Note that as shown in the Appendix, we have $\tilde{\alpha} \ll \alpha$ for α close to 0 and $(1 - \tilde{\alpha}) \ll (1 - \alpha)$ for α close to 1. Therefore, we find $w_\alpha := E(w_{i,\alpha}) = (1 - \tilde{\alpha})\alpha + \tilde{\alpha}(1 - \alpha) \approx \alpha$, so that with (2.4), we may approximate (2.6) through expansion to

$$\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}} \approx \lambda^{-1} \sum_{i=1}^{n} w_{i,\tilde{\alpha}}(Y_i - m_{\tilde{\alpha}}) - \tag{2.7}$$

$$\lambda^{-2} \sum_{i=1}^{n} (w_{i,\tilde{\alpha}} - w) \sum_{j=1}^{n} w_j(Y_j - m_{\tilde{\alpha}}) + \cdots \quad . \tag{2.8}$$

The first component in (2.7) has mean zero and variance

$$V_{\tilde{\alpha}} := \lambda^{-2}\Big[(1-\tilde{\alpha})^2\big\{H(q_\alpha) - 2G(q_\alpha)q_\alpha + q_\alpha^2\alpha\big\}$$

$$+ \tilde{\alpha}^2\big\{(H(\infty) - H(q_\alpha))(m_{0.5} - G(q_\alpha)) + (1-\alpha)q_\alpha^2\big\}\Big]. \qquad (2.9)$$

Note that with the assumption of finite second-order moments, we have $\int_{-\infty}^{q_\alpha} y^2 f(y)dy < \infty$ which implies that $f(y) = o(|y|^{-3})$ for $y \to -\infty$. This in turn yields that $\alpha q_\alpha^2 = o(1)$ and $G(q_\alpha)q_\alpha = o(1)$ so that overall $V_{\tilde{\alpha}} = o(1)$.

Looking at the second component in (2.8), we find that its mean equals $\lambda^{-2}\tilde{\alpha}(q_\alpha - m_{0.5})$ while its variance is of order $O(\tilde{\alpha}^2)O(V_\alpha)$. Arguing that $\tilde{\alpha} \ll \alpha$, see the Appendix, we can conclude that the second term in (2.8) is of ignorable asymptotic order for $\alpha = \lambda/n \to 0$. The same holds by simple calculation for the remaining components not explicitly listed in (2.7) and (2.8). Hence, we may approximate the distribution of $\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ through

$$\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}} \approx \lambda^{-1}\sum_{i=1}^{n} w_{i,\tilde{\alpha}}(y_i - m_{\tilde{\alpha}}). \qquad (2.10)$$

In particular, with (2.10), we get the (asymptotic) unbiasedness $E(\hat{m}_{\tilde{\alpha}}) = m_{\tilde{\alpha}}$. One may even derive asymptotic normality from (2.10) by showing that higher order moments vanish. We can, therefore, conclude that even for extreme expectiles, i.e., $\tilde{\alpha} = h(\alpha)$ or $1 - \tilde{\alpha} = 1 - h(\alpha)$ very small, respectively, we achieve asymptotic unbiasedness and normality.

We now pose the same question to quantiles, i.e., what can be said asymptotically about quantile estimation in the tails of the distribution. Following Koenker and Bassett (1978) and Koenker (2005, pp. 71–72), we can derive the distribution of the quantile estimate $\hat{q}_\alpha$ as follows. Let $g_\alpha(q) = \frac{1}{n}\sum_{i=1}^n 1\{Y_i \leq q\} - \alpha$ with $1\{.\}$ as a indicator function, then $\hat{q}_\alpha$ is defined through $g_\alpha(\hat{q}_\alpha) \geq 0$ and $g_\alpha(\hat{q}_\alpha - \delta) < 0$ for all $\delta > 0$. Hence

$$P(\hat{q}_\alpha - q_\alpha \leq \delta) = P\Big(\sum_{i=1}^n 1\{Y_i \leq q_\alpha + \delta\} \geq n\alpha\Big) = 1 - P(Z_\delta < n\alpha), \qquad (2.11)$$

where Z_δ is a binomial random variable with parameters $Z_\delta \sim Bin(n, F(q_\alpha + \delta))$. Note that $F(q_\alpha + \delta) \approx \alpha + f(q_\alpha)\delta$ and with (2.4) we have that the distribution of Z_δ (for small δ) converges to a Poisson distribution. As a consequence, for extreme quantiles, we do not achieve asymptotic normality and, therefore, unbiasedness is not guaranteed. We can easily calculate the limit of $P(\hat{q}_\alpha \leq q_\alpha)$ which equals $1 - P(Z \leq \lambda)$ for $Z \sim Poisson(\lambda)$. For instance, for $\lambda = 1$, this equals 0.26, which mirrors skewness of the distribution of extreme quantiles. Note, of course, that we may use extreme value theory to derive the asymptotic distribution of $\hat{q}_\alpha$.
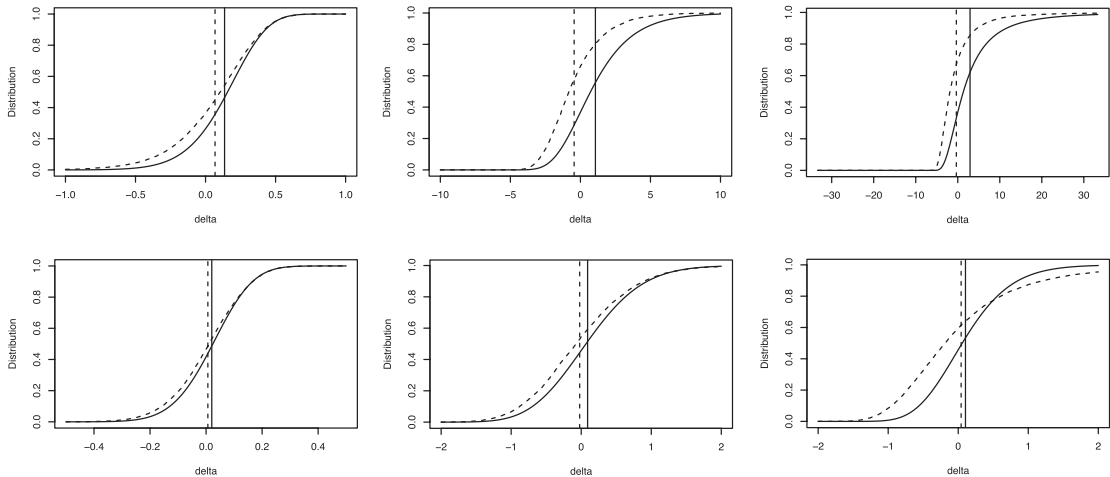
Figure 3 Sampling distribution of $\delta = \hat{q}_\alpha - q_\alpha$ (solid line) and $\delta = \hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ (dashed line) for different underlying distribution functions. Top row is for $\alpha = 0.999$ and bottom row for $\alpha = 0.99$. Left column is for normal distribution, middle column for Chi-squared (df = 2) distribution and right column for t-distribution (df = 3). The vertical line indicates the mean values for both distributions
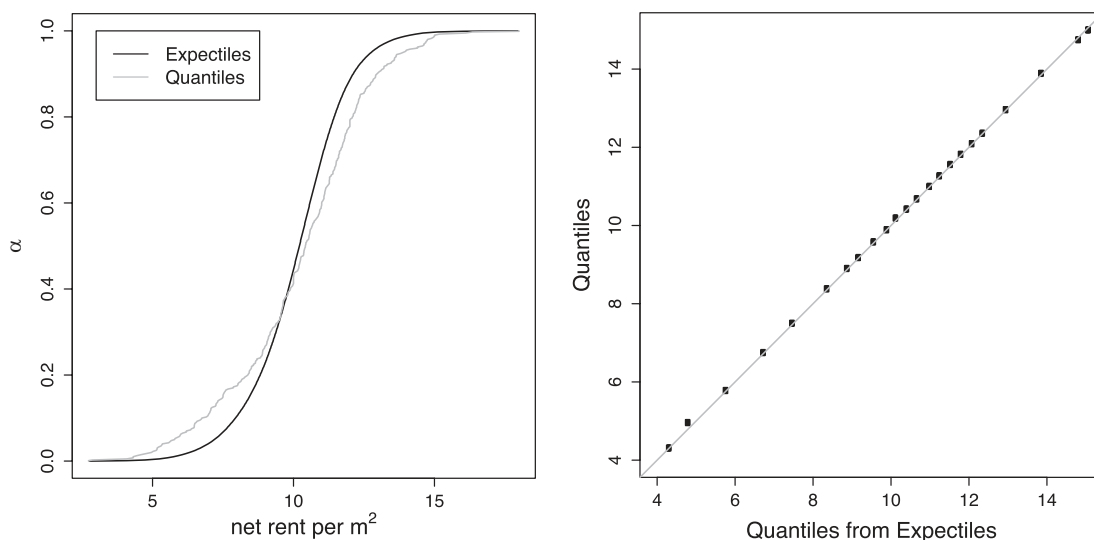**Source:** Authors' own (prepared with GNU R).

We run a small simulation to study the performance of tail expectile and tail quantile estimation. We simulate data and look at the distribution of extreme quantiles. In Figure 3 we show the distribution of $\hat{q}_\alpha - q_\alpha$ (solid line) and $\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ (dashed line) for a sample size of $n = 1000$. We look at the $\alpha = 0.999$ (top row) and the $\alpha = 0.99$ (bottom row) quantiles and the corresponding expectile. We simulate from (a) a normal distribution (left column), (b) a Chi-squared (with two degrees of freedom, middle column) and (c) a t-distribution (with three degrees of freedom, right-hand side column). The vertical line indicates the mean value of $\hat{q}_\alpha - q_\alpha$ (solid line) and $\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ (dashed line) which should be zero to indicate unbiasedness. There is apparently a bias occurring for quantile estimation for α close to 1 (or close to zero).

<u>**Conclusion:**</u> Overall we may conclude that expectiles estimates behave stable even for very small or very large values of α. This is of course important to know if one uses a sequence including even extreme expectiles to estimate quantiles as suggested in the previous subsection.

2.4 Example

To illustrate expectiles and the conversion of expectiles to quantiles we give a short example. We apply our methods to data collected 2012 in Munich to construct the Munich rent index. The full data set consists of 3080 observations, i.e., rented apartments in Munich, Germany, and we here analyze the variable giving the net rent per squared meter (m^2) for each apartment. For illustration, we restrict our attention to

Figure 4 Left: Quantiles and expectiles for the Munich rent data. Right: QQ-Plot quantiles versus quantiles from expectiles
Source: Authors' own (prepared with GNU R).

apartments between 45 m^2 and 55 m^2 and examine the net rent per square for the resulting 421 apartments in the data set of that size.

First we calculate a fine grid of sample expectiles and quantiles for the variable net rent per m^2 and plot them in Figure 4 on the left. The estimated expectiles naturally form a smooth curve while the estimated quantiles mirror some variability. In a next step, we use the set of expectiles to calculate quantiles from expectiles and plot these against the empirical quantiles (see the right part of Figure 4). When zooming into the figure, one can notice, that the estimated quantiles for the inner range of $\alpha \in (0, 1)$ nearly coincide with their empirical counterpart, whereas for extreme values of α there is some minor fluctuation around the identity line. This behaviour is also supported by our simulation results in Section 2.2. All in all, we see the applicability of the approaches. The example will be further discussed in subsection 4.2 by regressing the rent on the floorspace.

3   Quantile and expectile regression

3.1   The problem of crossing quantiles and expectiles

So far we have considered the simple scenario with no explanatory variables involved. We extend this now to quantile and expectile regression. To do so, we assume a continuous covariate x and define the quantile and expectile regression functions through $q_\alpha(x) = \beta_{0\alpha}^{(q)} + x\beta_{1\alpha}^{(q)}$ and $m_\alpha(x) = \beta_{0\alpha}^{(m)} + x\beta_{1\alpha}^{(m)}$. Estimation of $q_\alpha(x)$ and $m_\alpha(x)$

is carried out using the weighted L_1 sum (1.1) and the corresponding L_2 version (1.2), respectively, with $q_{i,\alpha} = q_\alpha(x_i)$ and $m_{i,\alpha} = m_\alpha(x_i)$.

A central problem occurring in quantile and expectile regression is crossing of fitted functions. For $0 < \alpha < \alpha' < 1$, by definition, we have $q_\alpha(x) < q_{\alpha'}(x)$ and $m_\alpha(x) < m_{\alpha'}(x)$, for all x. This inequality can, however, be violated for some (observed) x values in the fitted functions, which is called the crossing quantile or expectile problem. Several remedies have been proposed to circumvent the problem, some of which will be used later in the next section. Before turning to that, we want to explore empirically in simulations how frequently one is faced with crossing quantile and expectile functions. We run a small simulation study and count the number of crossings between neighbouring fitted expectiles and quantiles. Therefore, we select a set of $\alpha \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98, 0.99\}$ and simulate data from the following simple linear regression setting

$$y = 4 + 3x + \varepsilon. \tag{3.1}$$

The covariate x is drawn from a uniform $U(-1, 1)$ distribution, and the random error is added from either (a) a normal $N(0, 1.5^2)$, (b) a Chi-squared ($df = 2$) or (c) a t-distribution ($df = 3$). Data sets are generated with sample sizes of $n = 49, 199, 499$ and for each combination of settings 1000 replications are created. A data set is then analyzed by computing the set of α-quantiles using the R-package 'quantreg'. For expectiles, we compute the resulting $\tilde{\alpha} = h(\alpha)$ expectiles using the R-package 'expectreg'. Function $h(.)$ is computed separately for each error distribution according to (1.3), which makes the estimates comparable. For each data set and every generated covariate value, all neighbouring pairs of α and $h(\alpha)$ are checked for crossing regression lines within the range of observed covariates.

The resulting number of crossings within the 1000 replications is summarized in Table 1. Not surprisingly, crossings occur in the tail of the distribution and become less frequent with increasing sample size. However, the numbers show that there are generally fewer crossings of expectiles, in particular, within the central 90% of the distribution, while for quantile regression we obtain a large proportion of crossings for small samples even in the inner part of the distribution, i.e., between the 0.2 and 0.8 quantiles.

<u>Conclusion:</u> We may conclude from the simulation that expectiles seem less vulnerable for crossing problems than quantile estimates.

3.2 Non-crossing spline-based estimation

Several remedies have been suggested to circumvent or correct for crossing quantiles with references given in the introduction. We here extend the idea of Bondell *et al.* (2010) who fit non-crossing quantiles using linear programming. We pick up the idea and generalize it towards non-crossing spline-based expectile estimation. To do so, we first present spline-based quantile and expectile estimation by replacing

Table 1 Number of crossings between two neighbouring expectiles/quantiles in the linear model (3.1) from 1000 data sets, starting with $\alpha = 0.01$. Crossing counts are given for all 10 pairs of expectiles or quantiles, respectively, sample sizes of $n = 49, 199, 499$ and the three error distributions for ε, as defined previously. Quantiles smaller than $1/n$ are omitted and indicated as $*$ in the table

expectiles									
$\varepsilon \sim$	$N(0, 1.5^2)$			$\chi^2(2)$			$t(3)$		
$h(\alpha)$ with α / n	49	199	499	49	199	499	49	199	499
0.01–0.02	$*$	21	5	$*$	56	1	$*$	49	48
0.02–0.05	54	2	0	207	3	0	22	35	24
0.05–0.1	8	0	0	44	0	0	7	18	3
0.1–0.2	1	0	0	4	0	0	5	5	2
0.2–0.5	0	0	0	0	0	0	1	0	0
0.5–0.8	0	0	0	1	0	0	0	0	0
0.8–0.9	0	0	0	5	0	0	10	3	1
0.9–0.95	8	0	0	19	0	0	14	11	2
0.95–0.98	48	2	0	42	0	0	23	27	23
0.98–0.99	$*$	27	3	$*$	27	1	$*$	53	45

quantiles									
$\varepsilon \sim$	$N(0, 1.5^2)$			$\chi^2(2)$			$t(3)$		
α/n	49	199	499	49	199	499	49	199	499
0.01–0.02	$*$	564	210	$*$	593	226	$*$	612	296
0.02–0.05	714	123	10	758	143	16	679	208	27
0.05–0.1	443	46	0	439	46	0	377	78	3
0.1–0.2	144	6	0	167	3	0	156	2	0
0.2–0.5	5	0	0	5	0	0	3	0	0
0.5–0.8	3	0	0	13	0	0	4	0	0
0.8–0.9	168	2	0	154	5	0	166	7	0
0.9–0.95	433	37	1	432	40	1	411	78	3
0.95–0.98	680	130	12	718	175	18	704	214	37
0.98–0.99	$*$	593	213	$*$	599	199	$*$	647	267

Source: Authors' own (prepared with GNU R).

$q_\alpha(x)$ and $m_\alpha(x)$ with bivariate functions

$$q(\alpha, x) \text{ and } m(\alpha, x), \tag{3.2}$$

where both functions are smooth (or just linear) in the direction of α and smooth (or just linear) in the direction of x. The bivariate functions may be called *quantile sheets* or *expectile sheets*. The setting (3.2) transfers the estimation exercise to bivariate smoothing, as proposed in Schnabel and Eilers (2013b) and Schnabel and Eilers (2009a). We replace (or approximate) $q(\alpha, x)$ through

$$q(\alpha, x) = [\boldsymbol{B}^{(1)}(\alpha) \otimes \boldsymbol{B}^{(2)}(x)] \, \boldsymbol{u}, \tag{3.3}$$

where $\boldsymbol{B}^{(1)}(\alpha)$ is a (linear) B-spline basis set up on knots $0 < \alpha_1 < \cdots < \alpha_L < 1$, $\boldsymbol{B}^{(2)}(x)$ is a B-spline basis built upon some knots $\tau_1 < \cdots < \tau_K$ covering the range of observed

values of x and $\boldsymbol{u}$ is the vector of coefficients. If $q(\alpha, x)$ is assumed to be linear in x, one may take $\boldsymbol{B}^{(2)}(x)$ as linear B-spline and set $K = 2$ in this case. Let $l = 1, \ldots, L$ be the indices of columns of $\boldsymbol{B}^{(1)}(\alpha)$, and $k = 1, \ldots, K$ the indices of columns of $\boldsymbol{B}^{(2)}(x)$. Vector $\boldsymbol{u}$ may then be indexed by u_{lk} for $l = 1, \ldots, L$, $k = 1, \ldots, K$, and let $\boldsymbol{u}_{l.} = (u_{l1}, \ldots, u_{lk})^T$. Non-crossing quantiles are now guaranteed by linear constraints on the parameter vector of the form

$$\boldsymbol{B}^{(2)}(x)(\boldsymbol{u}_{l.} - \boldsymbol{u}_{l+1.}) \leq 0 \text{ for } l = 1, \ldots, L - 1 \tag{3.4}$$

for all x in the observed range of the covariates. If $\boldsymbol{B}^{(2)}(x)$ is a linear B-spline basis, this simplifies to $u_{lk} \leq u_{l+1k}$ for $l = 1, \ldots, L - 1$ and $k = 1, \ldots, K$. In general, we can formulate (3.4) as a linear constraint by inserting for x the observed values $x_i, i = 1, \ldots, n$.

We can now fit $q(\alpha, x)$ by replacing (1.1) with its multiple version

$$\sum_{l=1}^{L} \sum_{i=1}^{n} w_{i,\alpha_l} \left| y_i - \boldsymbol{B}^{(1)}(\alpha_l) \otimes \boldsymbol{B}^{(2)}(x_i) \, \boldsymbol{u} \right| \tag{3.5}$$

which is minimized with respect to $\boldsymbol{u}$ subject to the linear constraints (3.4) using linear programming. Alternatively, one may work with iterated weighted least squares by using the fact that $|y - q_\alpha| = (\sqrt{(y - q_\alpha)^2})^{-1}(y - q_\alpha)^2$. Schnabel and Eilers (2013b) change the weight from $w_{i,\alpha}$ to $w_{i,\alpha}(\sqrt{(y - q_\alpha)^2})^{-1}$ and apply iterated weighted least squares to fit the function $q(\alpha, x)$.

Replacing the L_1 distance in (3.5) by the L_2 distance

$$\sum_{l=1}^{L} \sum_{i=1}^{n} w_{i,\alpha_l} \left(y_i - \boldsymbol{B}^{(1)}(\alpha_l) \otimes \boldsymbol{B}^{(2)}(x_i) \, \boldsymbol{v} \right)^2 \tag{3.6}$$

gives a weighted least-squares criterion which allows to estimate the expectile sheet $m(\alpha, x) = \boldsymbol{B}^{(1)}(\alpha) \otimes \boldsymbol{B}^{(2)}(x) \, \boldsymbol{v}$, where again the linear constraints (3.4) need to be fulfilled. Estimation can be carried out by iterative quadratic programming.

In Figure 5, we show an example for the simulation model (3.1) of the previous subsection. We plotted the resulting quantile and expectile sheets under the assumption of normally distributed residuals. We use a linear B-spline basis for x with $K = 2$. For every value, we have increasing (or nondecreasing) functions $q(\alpha, x)$ and $m(\alpha, x)$ in α. A simple visual impression shows that the quantile sheet is more variable compared to the fitted expectile sheet. Note that we can now calculate for each value of x a set of expectiles $\hat{m}_{\alpha_l}(x) = \hat{m}(\alpha_l, x)$, which allows to apply the results of Section 2 to derive quantiles based on expectiles. The code for calculating the linear non-crossing quantile curves by Bondell *et al.* (2010) is available from the homepage of Howard Bondell (see `http://www4.stat.ncsu.edu/~bondell/Software/NoCross/`
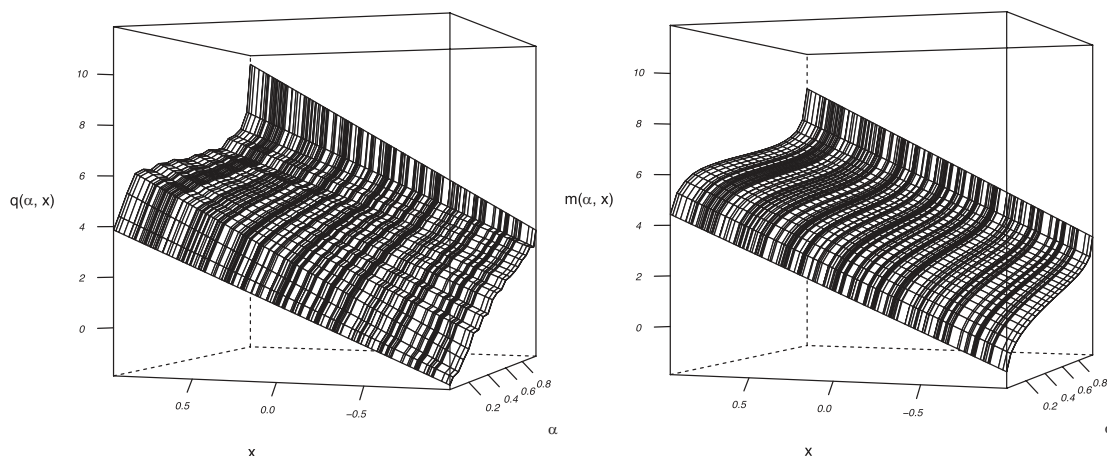
Figure 5 Quantile and expectile sheets for normally distributed errors
Source: Authors' own (prepared with GNU R).

`NoCrossQuant.R`, last date of access 10 January 2014). The programme for fitting non-crossing expectiles is part of the R package 'expectreg' by Sobotka *et al.* (2013).

<u>**Conclusion:**</u> For both expectiles and quantiles, we can fit sheets guaranteeing non-crossing functions. Overall, the expectile sheet provides a more smooth surface compared to the quantile sheet, in particular, in the direction of α.

4 Extensions and examples

4.1 Penalized smooth expectile sheets

Following the expectile sheet $m(\alpha, x)$, we may assume that $m(\alpha, x)$ is smooth in x, but without any parametric (linear) assumption. This can be fitted with a B-spline basis as in Section 3.2, but now with K being large. In order to control for a smooth and numerically stable fit, one may impose a penalty on the coefficients in the style of penalized spline regression (see Ruppert *et al.*, 2003, 2009). In other words we supplement (3.6) by the quadratic penalty

$$\lambda_v \boldsymbol{v}^T \boldsymbol{D}^T \boldsymbol{D} \boldsymbol{v},$$

where $\boldsymbol{D}^T \boldsymbol{D}$ is an appropriately chosen penalty matrix, and λ_v is the smoothing parameter which is chosen data driven. We give an example in the next subsection. This was first proposed by Schnabel and Eilers (2009a) with an additional penalization in direction of α. For a specific value of α this has been proposed in Sobotka *et al.* (2012) for expectile smoothing or in Bollaerts *et al.* (2006) for quantile smoothing where the latter use a different penalization. The smoothing parameter λ_v can be chosen by

asymmetric cross-validation or the Schall algorithm for mixed models as described in Schnabel and Eilers (2009b).

4.2 Rent index of Munich

To see how the method performs in practice, we again take a look at the Munich rental data from Section 2.4. As a reminder, the data consists of 3080 observations, i.e., rented apartments in Munich, Germany. We consider two variables in our example: net rent per m^2 as response and living space measured in m^2 as covariate. First, we perform both a non-crossing and non-parametric expectile regression as described in Sections 3.2 and 4.1, respectively. Our underlying model is given with the expectile sheet $m(\alpha, \text{living space})$. The smoothing parameter λ_v was chosen automatically by using the Schall algorithm (see Schnabel and Eilers, 2009b). The sheet resulting from the estimation procedure is shown in Figure 6. As one can see the dependency of the two variables is obviously of non-linear nature. The amount of smoothing done for the expectile sheet seems appropriate. The sheet serves as a basis to calculate quantile estimates for certain values of $x = 25, 30, \ldots, 155$ and $\alpha = 0.01, 0.02, 0.05, 0.10, 0.20, 0.50, 0.80, 0.90, 0.95, 0.98, 0.99$. We apply the algorithm as described in Section 2 and obtain the mid-panel of Figure 6. The calculated values for the quantiles are indicated by points which are connected by lines. All in all, the quantiles from expectiles seem to behave well. We can see that there is a decrease in net rent per square metre as the apartment size grows. This continues for apartments up to size 100 m^2, but then net rent remains, more or less, constant. A nice feature of our conversion is that non-crossing of quantiles is guaranteed.

As an alternative to the expectile-based analysis, we apply smooth non-crossing spline-based quantile fitting as described in Muggeo *et al.* (2013) and implemented in the R-package 'quantregGrowth'. Here, we choose cubic B-splines with a penalization through second-order differences. Muggeo *et al.* (2013) guarantee non-crossing of quantile curves by imposing inequality constraints on the spline coefficients. As the R-package allows us to estimate growth charts, there is also the possibility to enforce monotonicity in the direction of the dependent variable. The resulting fit is shown in the right part of Figure 6. All in all, spline-based quantiles exhibit similar features as the quantiles from expectiles, although the quantile smoothing spline, due to its L_1 nature, is angled. For the non-crossing quantile smoothing, we decided to pick a smoothing parameter which would result in a smoothness comparable to the amount of smoothing mirrored in the second panel of Figure 6.

4.3 Expected shortfall

Investment risks are frequently measured using the ES, a stochastic risk measure, for the lower tail defined as $ES(\alpha) = E(Y|Y < q_\alpha)$ for a continuous random variable Y with α-quantile q_α. It measures the expectation given that the random variable does not exceed a fixed value and is often applied to financial time series. A naïve estimate would calculate the mean beyond a previously estimated quantile and would,
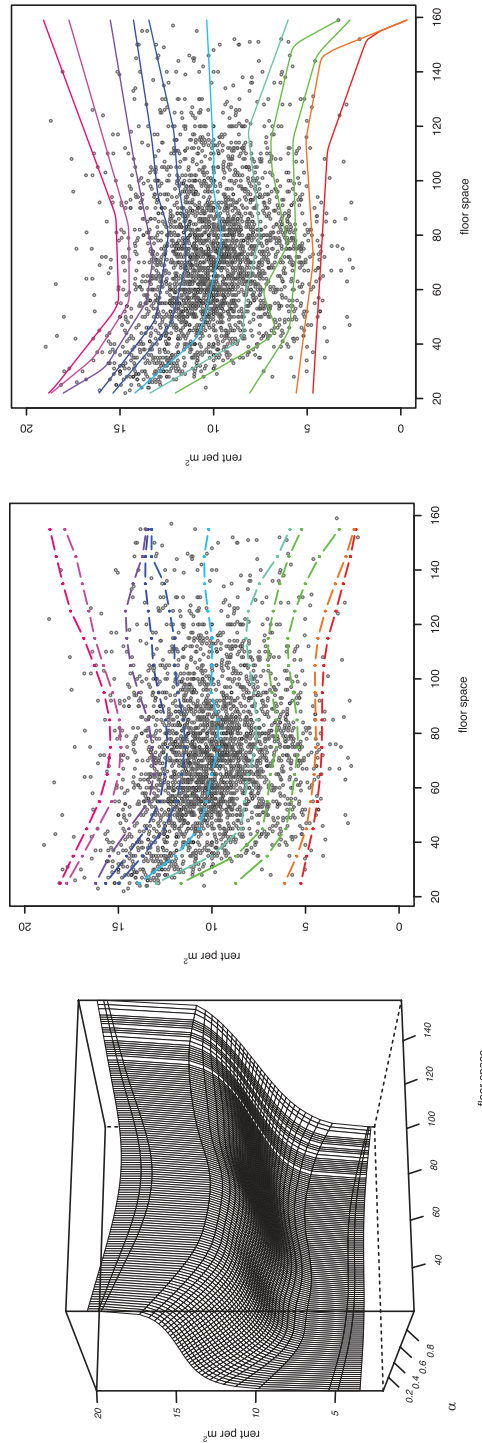
Figure 6 Expectile sheet for the Munich rent data (left), quantiles based on expectiles, calculated for certain values of x, which are visualized by points (middle) and non-crossing quantile smoothing splines (right). Quantiles are calculated for α = 0.01, 0.02, 0.05, 0.10, 0.20, 0.50, 0.80, 0.90, 0.95, 0.98, 0.99

Source: Authors' own (prepared with GNU R).

therefore, be rather inefficient. Taylor (2008) presents a possibility to estimate the ES using expectiles and their connection to quantiles.

Note that the $\tilde{\alpha}$-expectile is implicitly defined through $\arg\min E(w_{i,\tilde{\alpha}}(y_i - m)^2)$ so that the expectile satisfies

$$\frac{1 - 2\tilde{\alpha}}{\tilde{\alpha}} E[(Y - m_{\tilde{\alpha}})I(Y < m_{\tilde{\alpha}})] = m_{\tilde{\alpha}} - E(Y), \tag{4.1}$$

where, as above, $\tilde{\alpha} = h(\alpha)$. That is, the expectile $m_{\tilde{\alpha}}$ is determined by the expectation of the random variable Y conditional on $Y < m_{\tilde{\alpha}}$. Rewriting (4.1) and using the fact $F(m_{\tilde{\alpha}}) = \alpha$ leads to

$$ES_{\text{low}}(\alpha) := E(Y|Y < q_\alpha) = \left(1 + \frac{\tilde{\alpha}}{(1 - 2\tilde{\alpha})\alpha}\right)m_{\tilde{\alpha}} - \frac{\tilde{\alpha}}{(1 - 2\tilde{\alpha})\alpha}m_{0.5} \tag{4.2}$$

for the lower tail of F. Depending on whether the random variable describes a win or a loss, we define the ES for the upper tail as $ES_{\text{up}}(\alpha) = E(Y|Y > m_{1-\tilde{\alpha}})$. In order to determine the appropriate $\tilde{\alpha}$ to a given α, Taylor (2008) estimates a dense set of expectiles and then constructs an empirical distribution function on the basis of the expectile curves. Here, we make use of the results derived in subsection 2.1 and estimate the distribution function $\hat{F}_m(.)$ from expectiles. As introduced in Section 2.1, we estimate a dense set of expectiles (i.e., we set $\alpha_l = 0.0005, 0.001, 0.005, 0.01, 0.02, \ldots, 0.98, 0.99, 0.995, 0.999, 0.9995$) and compute the cumulative distribution function at each observed covariate value. The estimated distribution allows us to conclude the $\tilde{\alpha}$ value for a given α-quantile. We then calculate the ES explicitly for certain values of α, e.g., $\alpha = 0.01, 0.05, 0.95, 0.99$.

We apply the idea and estimate the ES for a serially drawn time series from the daily yields of the French stock index CAC40 in the time period between 1991 and 1998. All in all, there are 1860 observations/trading days and we take time t as the covariate influencing the ES. For estimation, we, therefore, construct the expectile sheet $m(\alpha, t)$. As basis in t, we use a cubic B-spline basis with 20 inner knots to account for the variability in time. However, in order to give a risk prediction for the next observations, we have equidistant knots from $\min(t)$ to $\max(t) + 0.02(\max(t) - \min(t))$. That way, we get an estimated risk for the upcoming day(s), i.e., we pursue out of range prediction. To achieve smoothness in time, we add a penalty of first-order differences $\lambda_v v^T D^T D v$ to (3.6), where Dv has rows $v_{lk} - v_{lk-1}$ for $l = 1, \ldots, L$ and $k = 1, \ldots, K$. The optimal smoothing parameter λ_v is chosen via asymmetric cross-validation; see Sobotka *et al.* (2012) for a more extensive description. Next, we apply the algorithm presented in Section 2 to all observed covariate values and also to the added time points beyond the data. This delivers the estimated α-quantile and its corresponding $\tilde{\alpha}$-expectile. In turn we are able to estimate the expected shortfall (4.2) for each point in time. This is done sequentially, that is, observation by observation, to gather information about the changes in the distribution.

The result of the estimation is presented in Figure 7 (left part). For comparison, we also fit the ES based on the empirical distribution as suggested by Taylor (2008)
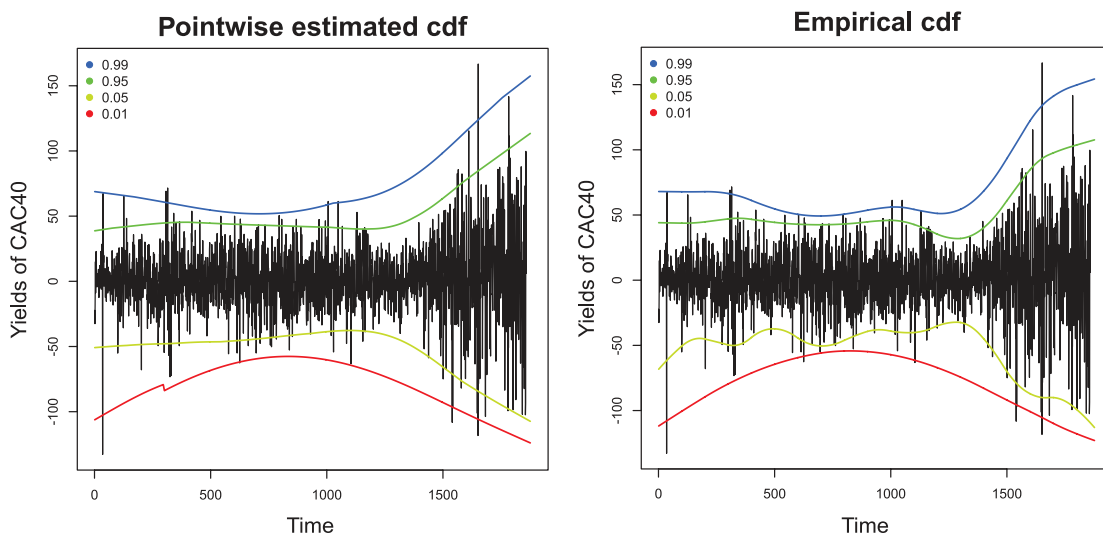
Figure 7 Estimated ES of the CAC40 yields for $\alpha = 0.01, 0.05, 0.95, 0.99$. Daily data from 1991 to 1998. Results with pointwise estimated distribution on the left, and with empirical distribution function on the right
Source: Authors' own (prepared with GNU R).

(the right part of Figure 7). As can be seen, the volatility of the data is captured by the curves of the ES, for gains, as well as for losses. A generalization over the range of time can also be observed. When using the empirical distribution function on the other hand, especially the curves for $\alpha = 0.05, 0.95$ tend towards overfitting. This is particularly visible for the time of low volatility around day 1300. The small amount of prediction incorporated by the splines turns out to be just a linear extension of the last fits. For accurate predictions, one should aim to combine conditional autoregressive expectiles (CARE, Kuan *et al.*, 2009) that are able to account for the autocorrelation in the data with the methods introduced in this article. Still, the example shows that an improvement in ES estimation is possible when using the efficient distribution estimation introduced in Section 2.1.

5   Discussion

In this article, we looked at quantiles, as Goliath, and expectiles, as David, and explored how their connection can be used in practice. An algorithm was presented to estimate quantiles from a (fine) grid of expectiles. We noticed and examined properties of extreme quantiles and expectiles and discussed the crossing issue of quantile and expectile regression. Even so, as crossing of neighbouring curves is an issue, we proposed a modified method to circumvent this problem. All methods regarding expectiles which were described in detail in this article can be found in the R-package 'expectreg'.

Apparently, referring again to the comparison of expectiles and quantiles to David and Goliath is undissolved. There is no final fight, and research on both

ends continues. It is certainly true that quantiles are dominant in the literature but we wanted to show that expectiles are an interesting alternative to quantiles and that their combined use is helpful, in particular, for the estimation of the ES. We also demonstrated the use of quantile and expectile sheets as smooth variants to quantile and expectile regression, respectively. This accommodates quite naturally the constraints of non-crossing quantile and expectile curves and the latter allows for smooth expectile regression based on implemented software, as mentioned above. Also, expectile regression now can be performed without loosing interpretability, since quantiles can be estimated from expectiles.

All in all, we hope to have convinced the reader that expectiles do not immediately 'belong in the spittoon' as Koenker (2013a) provocatively postulates. We think that expectiles provide an interesting and worthwhile alternative to the well-established quantile regression.

Acknowledgements

Appendix

A.1 Relation of α and $\tilde{\alpha}$ for $\alpha \to 0$

We assume that y has finite second moments, then with (1.3)

$$\frac{\alpha}{\tilde{\alpha}} = \frac{-\alpha m_{0.5} + 2\alpha G(q_\alpha) + \alpha(1 - 2\alpha)q_\alpha}{-\alpha q_\alpha + G(q_\alpha)}. \tag{A.1}$$

Since the nominator and denominator both tend to zero for $\alpha \to 0$, we apply the rule of de l'Hospital. Observing that $\alpha q_\alpha = o(1)$ for $\alpha \to 0$, we get

$$\lim_{\alpha \to 0} \frac{\alpha}{\tilde{\alpha}} = \lim_{\alpha \to 0} -\frac{f(q_\alpha)(q_\alpha - m_{0.5}) + \alpha}{\alpha} > 1. \tag{A.2}$$

Hence $\tilde{\alpha} < \alpha$ for $\alpha \to 0$. Note that since $f(q_\alpha) = o(|q_\alpha|^{-3})$ which follows due to the existence of second-order moments, we find again that nominator and denominator of (A.2) converge to zero for $\alpha \to 0$. Assuming now that $f(q_\alpha)$ is proportional to $|q_\alpha|^{-3}$ for $\alpha \to 0$ which is required to guarantee finite second-order moments, we get, again with the rule of de l'Hospital applied to (A.2), that $\lim_{\alpha \to 0} \alpha/\tilde{\alpha} = const \geq 1$, while if $f(q_\alpha)$ is proportional to $|q_\alpha|^{-(3+\delta)}$ for some $\delta > 0$, we get with the same arguments that $\lim_{\alpha \to 0} \alpha/\tilde{\alpha} = \infty$.

A.2 Estimation of $\hat{F}_m(.)$

Let $0 < \alpha_1 < \cdots < \alpha_L < 1$ be a dense set of knots covering $(0, 1)$ and containing 0.5 and define with l_0 the index with $\alpha_{l_0} = 0.5$. First note, that (2.1) for α_{l_0} gives a redundant information as it states that $m_{0.5} = m_{0.5}$. That is to say that we need an additional constraint. This is found by observing that

$$m_{0.5} = \int_{-\infty}^{\infty} y\, dF(y) = \frac{m_L + m_{L+1}}{2} + \sum_{l=1}^{L} (c_l - c_{L+1})\gamma_l \qquad \text{(A.3)}$$

with the approximation of $F(.)$ from Section 2.1. Remembering the definition of the expectiles (2.1), we define function $g_l(.)$ by

$$g_l(\gamma_l, \ldots, \gamma_1) = m_l - \frac{(1 - \alpha_l)G_l(\gamma_l, \ldots, \gamma_1) + \alpha_l(m_{0.5} - G_l(\gamma_l, \ldots, \gamma_1))}{(1 - \alpha_l)F_l(\gamma_l, \ldots, \gamma_1) + \alpha_l(1 - F_l(\gamma_l, \ldots, \gamma_1))} \quad \text{for}$$
$$l = 1, \ldots, L. \qquad \text{(A.4)}$$

We now need $\gamma_1, \ldots, \gamma_L$ such that $g_l \equiv 0$ which in principle can be seen as a root finding problem. We implemented a version where we minimize the sum of squares of $g_l(\gamma)$ under certain restrictions: We face the minimization problem

$$\min_{\gamma_1, \ldots, \gamma_L} S(\gamma_1, \ldots, \gamma_L) = \min_{\gamma_1, \ldots, \gamma_L} \sum_{l=1}^{L} (g_l(\gamma_l, \ldots, \gamma_1)^2) \qquad \text{(A.5)}$$

under the constraints that $\gamma_l \geq 0$ and $\sum_{l=1}^{L} \gamma_l \leq 1$ which is solved by Newton's method in optimization and also implemented in the R-package 'expectreg' by Sobotka *et al.* (2013). Penalty parameter λ_{pen}, which ensures numerical stability and smoothness of the distribution function, may be set equal to the squared empirical variance of the data from which the expectiles are estimated. In our simulations, we set λ_{pen} equal to five times the squared empirical variance of the data (for each of the three distributions considered and for both sample sizes $n = 199$ and $n = 499$).

References

Aigner DJ, Amemiya T and Poirier DJ (1976) On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, **17** (2), 377–96.

Andriyana Y, Gijbels I and Verhasselt A (2014) P-splines quantile regression estimation in varying coefficient models. *TEST*, **23**(1), 153–94.

Bollaerts K, Eilers P and Aerts M (2006) Quantile regression with monotonicity restrictions using p-splines and the l1-norm. *Statistical Modelling*, **6**(3), 189–207.

Bondell H, Reich B and Wang H (2010) Non-crossing quantile regression curve estimation. *Biometrika*, **97**, 825–38.

Breckling J and Chambers R (1988) M-quantiles. *Biometrika*, **75**, 761–71.

Chernozhukov V, Fernández-Val I and Galichon A (2010) Quantile and probability curves without crossing. *Econometrica*, **78**(3), 1093–125.

De Rossi G and Harvey A (2009) Quantiles, expectiles and splines. *Journal of Econometrics*, Nonparametric and Robust Methods in Econometrics (Special issue). **152**(2), 179–85.

Dette H and Scheder R (2011) Estimation of additive quantile regression. *Annals of the Institute of Statistical Mathematics*, **63**(2), 245–65.

Efron B (1991) Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, **1**, 93–125.

Eilers PHC (2013) Discussion: The beauty of expectiles. *Statistical Modelling*, **13**(4), 317–22.

Guo M and Härdle W (2013) Simultaneous confidence bands for expectile functions. *AStA – Advances in Statistical Analysis*, **96**(4), 517–41.

He X (1997) Quantile curves without crossing. *The American Statistician*, **51**(2), 186–92.

Honda T (2004) Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, **121**(1), 113–25.

Jones M (1992) Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics*, **44**(4), 721–27.

Jones M (1994) Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters*, **20**(2), 149–53.

Kneib T (2013) Beyond mean regression (with discussion and rejoinder). *Statistical Modelling*, **13**(4), 275–303.

Koenker R (1984) A note on l-estimates for linear models. *Statistics and Probability Letters*, **2**(6), 323–25.

Koenker R (2005) *Quantile regression, econometric society monographs*. Cambridge: Cambridge University Press.

Koenker R (2013a) Discussion of 'beyond mean regression' by T. Kneib. *Statistical Modelling*, **13**(4), 323–33.

Koenker R (2013b) *Quantreg: quantile regression*. R package version 4.97.

Koenker R and Bassett G (1978) Regression quantiles. *Econometrica*, **46**(1), 33–50.

Koenker R, Ng P and Portnoy S (1994) Quantile smoothing splines. *Biometrika*, **81**(4), 673–80.

Kuan CM, Yeh JH and Hsu YC (2009) Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, **150**(2), 261–70.

Leorato S, Peracchi F and Tanase AV (2012) Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics and Data Analysis*, **56**(4), 768–84.

Muggeo V, Sciandra M, Tomasello A and Calvo S (2013) Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environmental and Ecological Statistics*, **20**, 519–31.

Neocleous T and Portnoy S (2008) On monotonicity of regression quantile functions. *Statistics and Probability Letters*, **78**(10), 1226–29.

Newey WK and Powell JL (1987) Asymmetric least squares estimation and testing. *Econometrica*, **55**(4), 819–47.

Noh H, Chung K and Van Keilegom I (2012) Variable selection of varying coefficient models in quantile regression. *Electronic Journal of Statistics*, **6**, 1220–38.

Pratesi M, Ranalli M and Salvati N (2009) Nonparametric M-quantile regression using penalised splines. *Journal of Nonparametric Statistics*, **21**(3), 287–304.

Reiss P and Huang L (2012) Smoothness selection for penalized quantile regression splines. *International Journal of Biostatistics*, **8**(1).

Ruppert D, Wand MP and Carroll RJ (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.

Ruppert D, Wand MP and Carroll RJ (2009) Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*, **3**, 1193–256.

Schnabel SK and Eilers PHC (2009a, July 20–24) Non-crossing smooth expectile curves.

In JG Booth, ed. *Proceedings of the 24th International Workshop on Statistical Modelling*. Ithaca, NY, USA, 330–36.

Schnabel SK and Eilers PHC (2009b) Optimal expectile smoothing. *Computational Statistics and Data Analysis*, **53**(12), 4168–77.

Schnabel SK and Eilers PHC (2013a) A location-scale model for non-crossing expectile curves. *Stat*, **2**(1), 171–83.

Schnabel SK and Eilers PHC (2013b) Simultaneous estimation of quantile curves using quantile sheets. *Advances in Statistical Analysis*, **97**(1), 77–87.

Sobotka F, Kauermann G, Waltrup, L Schulze and Kneib T (2012) On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, **23**(2), 135–48.

Sobotka F and Kneib T (2012) Geoadditive expectile regression. *Computational Statistics and Data Analysis*, **56**(4), 755–67.

Sobotka F, Schnabel S and Waltrup L Schulze (2013) *Expectreg: expectile and quantile regression*. With contributions from P. Eilers, T. Kneib and G. Kauermann, R package version 0.39.

Taylor J (2008) Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, **6**(2), 231–52.

Wang H and Zhou X (2010) Estimation of the retransformed conditional mean in health care cost studies. *Biometrika*, **97**, 147–58.

Wu Y and Liu Y (2009) Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and Its Interface*, **2**, 299–310.

Yao Q and Tong H (1996) Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics*, **6**(2–3), 273–92.

Yuan M (2006) Gacv for quantile smoothing splines. *Computational Statistics and Data Analysis*, **50**, 813–29.

Ziegel JF (2013) Coherence and elicitability. arXiv:1303.1690v2.