Linda Schulze Waltrup

Extensions of Semiparametric Expectile Regression

Dissertation submitted to the Faculty of Mathematics, Informatics, and Statistics of the LMU Munich

Extensions of Semiparametric Expectile Regression

Linda Schulze Waltrup

Dissertation an der
Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

02.12.2014

Erster Berichterstatter:
Prof. Dr. Göran Kauermann (LMU München)

Zweiter Berichterstatter:
Prof. Dr. Thomas Kneib (Universität Göttingen)

Disputation: 11.2.2015

Zusammenfassung

Die Expektilregression kann als eine Erweiterung der Mittelwertsregression gesehen werden, da sie allgemeinere Eigenschaften der Verteilung einer interessierenden Größe beschreibt. Diese Arbeit führt in die Expektilregression ein und präsentiert neue Erweiterungen bereits existierender semiparametrischer Regressionsmodelle.

Diese Dissertation besteht aus vier zentralen Teilen. Als Erstes wird jeweils die bijektive Beziehung zwischen Expektilen und der empirischen Verteilungsfunktion beziehungsweise Quantilen genutzt, um aus einer großen Menge dicht beieinanderliegender Expektile die Verteilungsfunktion und Quantile zu berechnen. Sogenannte „Quantiles-from-expectiles"-Schätzer werden eingeführt und mit direkten Quantilschätzern in Bezug auf Effizienz verglichen. Als Zweites wird eine Methode zur Schätzung von nicht-kreuzenden Expektilkurven entwickelt. Zudem wird der Fall betrachtet, dass man geclusterte oder longitudinale Beobachtungen vorliegen hat. Hierfür wird eine zufällige, individuelle Komponente eingeführt, was zu einer Erweiterung der gemischten Modelle hin zu gemischten Expektilmodellen führt. Als Drittes wird eine Möglichkeit vorgestellt, die „Quantiles-from-expectiles"-Schätzer im Rahmen designbasierter Stichprobenverfahren zu schätzen.

Alle Methoden sind in der Open-Source-Software R implementiert und in einem R-Paket namens `expectreg` verfügbar. Eine Beschreibung des Paketes `expectreg` ist im vierten Teil dieser Arbeit enthalten.

Extensions of Semiparametric Expectile Regression

Linda Schulze Waltrup

Dissertation submitted to
the Faculty of Mathematics, Informatics, and Statistics
of the LMU Munich

02.12.2014

First referee: Prof. Dr. Göran Kauermann (LMU Munich)
Second referee: Prof. Dr. Thomas Kneib (Universität Göttingen)
Defense: 11.2.2015

Abstract

Expectile regression can be seen as an extension of available (mean) regression models as it describes more general properties of the response distribution. This thesis introduces to expectile regression and presents new extensions of existing semiparametric regression models.

The dissertation consists of four central parts. First, the one-to-one-connection between expectiles, the cumulative distribution function (cdf) and quantiles is used to calculate the cdf and quantiles from a fine grid of expectiles. Quantiles-from-expectiles-estimates are introduced and compared with direct quantile estimates regarding efficiency. Second, a method to estimate non-crossing expectile curves based on splines is developed. Also, the case of clustered or longitudinal observations is handled by introducing random individual components which leads to an extension of mixed models to mixed expectile models. Third, quantiles-from-expectiles-estimates in the framework of unequal probability sampling are proposed.

All methods are implemented and available within the package `expectreg` via the open source software R. As fourth part, a description of the package `expectreg` is given at the end of this thesis.

Danksagung

Ganz herzlich möchte ich mich bei meinem Doktorvater und Betreuer Göran Kauermann bedanken. Er hatte es während meines Studiums der Wirtschaftsmathematik an der Universität Bielefeld mit spannenden Vorlesungen im Bereich Statistik geschafft, mein Interesse für das Themengebiet zu wecken und dafür gesorgt, dass dieses Interesse auch während der Promotion nicht nachließ.

Thomas Kneib bin ich dankbar, dass er die Rolle des Zweitgutachters übernahm und als Ansprechpartner stets zur Verfügung stand.

Fabian Sobotka danke ich für unzählige Diskussionen über Expektilregression, für seine Hilfe in Sachen Programmierung und nicht zuletzt auch für die freundschaftliche Zusammenarbeit insgesamt.

Diese Arbeit entstand in großen Teilen während meiner Zeit am Institut für Statistik der LMU München. Die Anfänge dafür wurden jedoch am Lehrstuhl für Statistik an der Universität Bielefeld gelegt. In beiden Fällen hatte ich das Glück, wunderbare Kollegen zu haben, mit denen man nicht nur fachsimpeln, sondern auch immens viel Spaß haben konnte. Da wären zum einen meine Bürokolleginnen Nadja Kaufmann in Bielefeld und Stephanie Thiemichen in München zu erwähnen und zum anderen der harte Kern unserer Münchner Mittagsrunde bestehend aus Tina Feilke, Gunther Schauberger und Verena Maier. Die angenehme Atmosphäre am Münchner Institut hat mir den Wechsel von der beschaulichen Stadt Bielefeld hin in die süddeutsche Metropole München sicherlich sehr erleichtert. Auch außerhalb des Instituts habe ich in München gute Freunde gefunden, für die ich sehr dankbar bin!

Ablenkung habe ich auch bei meinen alten Schulfreunden und Kommilitonen erhalten. Ihnen bin ich dankbar für die manchmal dringend notwendige Zerstreuung!

Ganz besonders danken möchte ich meinen Eltern Birgit und Dietmar Schulze Waltrup, sowie meinen Großeltern Anni und Ferdi Schrage und Jenny und Wilhelm Schulze Waltrup. Meine Familie hat mich immer unterstützt und mir großen Halt gegeben. Dies gilt auch für meine Geschwister Maike und Robin Schulze Waltrup.

Seit vielen Jahren sehr viel Geduld mit mir hat Torben Teismann. Er stand mir bereits bei meinem ersten Abschluss, dem Abitur, beiseite und hat mich während des Studiums und der Promotion unzählige Male aufgemuntert und stets an mich geglaubt. Ihm danke ich für seine guten Nerven und dafür, dass er mich in allen Entscheidungen unterstützt hat.

München, im Dezember 2014 Linda Schulze Waltrup

Contents

Chapter 1

Introduction

1.1 Motivation

In order to understand relationships between variables it is often useful to model these relationships in a mathematical or, more precisely said, in a statistical way. Regression tries to explain the behaviour of one variable by imposing a dependency on one ore more (explanatory) variables. A common and simple procedure is to assume linear dependency and to model the variable of interest y as a linear function of the explanatory variable x (or $x_1, \dots, x_k$ if there is more than one). This results in classical linear regression. Unfortunately (or fortunately) a linear dependency is not always present in nature. Therefore, to give a good description of the data, it may be necessary to allow for nonlinear behaviour resulting in polynomial regression or, even more flexible, semi- or nonparametric regression. Both y and x can be seen as realisations of random variables Y and X, respectively. Often, however, the covariate x is regarded as fixed. In general the idea in nonparametric regression is to model y as a function of x which is only influenced by a so called error term ϵ, where ϵ represents random noise. That is we impose $y = f(x) + \epsilon$. Non- and semiparametric regression models are described in detail for example by Hastie and Tibshirani (1990), Ruppert, Wand, and Carroll (2003), Wood (2006), and Fahrmeir, Kneib, Lang, and Marx (2013).

Within the classical regression framework the expectation of the response y is modeled. In some cases other approaches may appear more reasonable, namely when especially high or low features of the data are of interest. Consider for example malnutrition of children (see Fenske, Kneib, and Hothorn, 2011) or frontier estimation as in Schnabel and Eilers (2009a). The former use quantile regression, the latter expectile regression to model the noncentral parts of a response distribution. In general there are several possibilities to estimate models which do not concentrate on the mean but on other (and often more) features of the data. A good overview of models "beyond mean regression" can be found in the correspondingly named paper by Kneib (2013). In this thesis we will concentrate on expectiles and expectile regression.

1.2 Outline

This thesis consists of five chapters based on four manuscripts. Chapter 2 gives an introduction to (semiparametric) regression and expectile regression. The content of Chapters 3 to 5 basically consists of three methodological papers where new extensions of semiparametric expectile regression are developed. Chapter 6 describes the implementation of the theoretical concepts.

Chapter 2 provides the theoretical background for readers which are not familiar with regression. The chapter starts with an introduction to linear and polynomial regression. After describing B-splines, we switch to semiparametric regression and semiparametric expectile regression. The interpretability of expectiles is also discussed at the end of the chapter.

Chapter 3 compares quantiles and expectiles concerning efficiency. A method to calculate quantiles from expectiles is presented to allow for a comparison on an equal level. This connection between quantiles and expectiles becomes extremely useful when calculating the expected shortfall. In addition, non-crossing expectile curves for a single covariate are developed. To demonstrate the usage, all procedures are accompanied by applications. Note that this chapter is developed in joint work with Fabian Sobotka, Thomas Kneib, and Göran Kauermann. A similar version will be published in Statistical Modelling under the title *Expectile and Quantile Regression - David and Goliath?*. All authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript. For a detailed description and assignment please look at the first page of Chapter 3.

In Chapter 4 non-crossing spline-based expectile estimates within an additive framework are developed. The use of a tensor product is encouraged to get expectile sheets. We also consider panel data, where we have longitudinal observations and extend the expectile regression model to allow for a random intercept. Chapter 4 concludes with an application of the newly developed estimators on data of the German Socio-Economic Panel (GSOEP). This chapter is based on a manuscript developed jointly with Göran Kauermann. Both Göran Kauermann and Linda Schulze Waltrup contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript. For a detailed description of the division of work please look at the first page of Chapter 4.

In Chapter 5 the estimation of quantiles from a set of expectiles is extended to samples where elements are drawn with unequal probabilities. A simulation is presented to compare the estimator with existing estimators in terms of bias and variance. This chapter is based on a manuscript developed jointly with Göran Kauermann. Both Göran Kauermann and Linda Schulze Waltrup contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript. For a detailed description and assignment please look at the first page of Chapter 5.

Chapter 6 is about the implementation of methods concerning expectiles. The estimation techniques described in the previous chapters are implemented in the open source software R (see R Core Team, 2014) within the R package `expectreg`. Chapter 6 gives descriptions and examples of use concerning the software. It bases on a manuscript developed jointly with Fabian Sobotka, Sabine Schnabel, Göran Kauermann, and Thomas Kneib and a former version of the manuscript is already published as part of the thesis *Semiparametric Expectile*

Regression by Fabian Sobotka, who is the leading author of the paper. All authors contributed to the general investigation of the underlying scientific problems and were involved in writing and proofreading the manuscript. For a more detailed description of the division of work please look at the first page of Chapter 6.

Chapter 7 summarizes the findings of the previous sections. The chapter ends with a short discussion and notes on future work in the field of expectile regression.

The Appendix at the end of this dissertation includes a compendium of the notation used during this dissertation (see Chapter A), and additional information regarding calculations of Chapter 3 (see Chapter B). For the sake of completeness there is also a fifth manuscript appended (see Chapter C). It is appeared in Statistics and Computing under the title *On confidence intervals for semiparametric expectile regression*. In this paper asymptotic results for the construction of confidence intervals are derived and evaluated empirically. The paper is developed by Fabian Sobotka, Göran Kauermann, and Thomas Kneib, and in minor parts by Linda Schulze Waltrup. All authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript. For a more detailed description of the division of work please look at the first page of Chapter C.

Chapter 2

Semiparametric Regression and Expectile Regression

2.1 Linear and Polynomial Regression

When linear dependency between y and x is a plausible assumption, the use of a linear regression is perfectly fine. Of course, one should also check other assumptions which are made within the regression framework (and which will be described later). Often, however, a linear model does not adequately describe the data. A first generalization would be to extend the classical linear model to a polynomial regression model. In the following, a short introduction into linear and polynomial regression is given. There are many textbooks explaining linear and polynomial regression and its extensions and generalizations in detail (see for example Hastie and Tibshirani, 1990; Ruppert, Wand, and Carroll, 2003; or Fahrmeir, Kneib, Lang, and Marx, 2013).

Suppose we have a response y, which we want to explain by covariate x. An example which is often used is the rent index of Munich where y is the net rent of an apartment. One possible explanatory variable x is the size of the apartment. If there are n apartments observed, one gets n realizations for y and x each, i.e. we have vectors $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{x} = (x_1, \ldots, x_n)'$. We restrict our attention to metrical variables in the following and variables x_i will be seen as fixed. The main idea in regression is to regard y as a function of x where exact measurement is only disturbed by a random noise ϵ, that is we have $y = f(x) + \epsilon$ with $\mathrm{E}(\epsilon_i) = 0$ for all $i = 1, \ldots, n$. We assume that for the errors ϵ it holds that they are identically and independently distributed with same variance $\mathrm{Var}(\epsilon_i) = \sigma^2$ for all $i = 1, \ldots, n$. The standard assumption in simple linear regression is, that we have a linear function f such that

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{for all} \quad i = 1, \ldots, n. \tag{2.1}$$

As the parameters β_0 and β_1 are unknown, we need to estimate them. There are several methods to estimate the unknown parameters like ordinary least squares (OLS), Bayes or Maximum-Likelihood. Descriptions of all methods can be found in the textbooks mentioned above. We will make use of OLS in the following. Defining the design matrix $\boldsymbol{X}$ and vectors $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ as

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

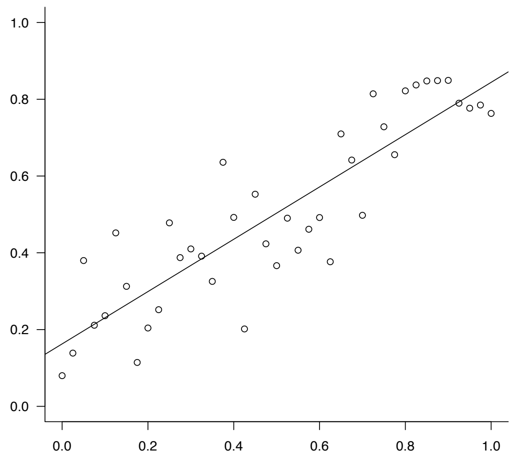


Figure 2.1: Linear model.

we get the linear model in matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.2}$$

OLS means minimizing squared differences. That is we obtain the estimator $\hat{\boldsymbol{\beta}}$ by minimizing

$$(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})' (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$. Differentiating and setting the first derivative equal to zero gives the normal equation which can be solved for $\boldsymbol{\beta}$. The solution is given as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y} \tag{2.3}$$

and we also get an estimation of $\mathrm{E}(\boldsymbol{y})$ calculating

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{y}, \tag{2.4}$$

where $\boldsymbol{H} := \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'$. Figure 2.1 shows an example of (simulated) data with a distinct linear trend. The Figure contains a scatterplot of the data with the straight line resulting from $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. Coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated from equation (2.3). The difference $y_i - \hat{y}_i$ is called residuum or residual.

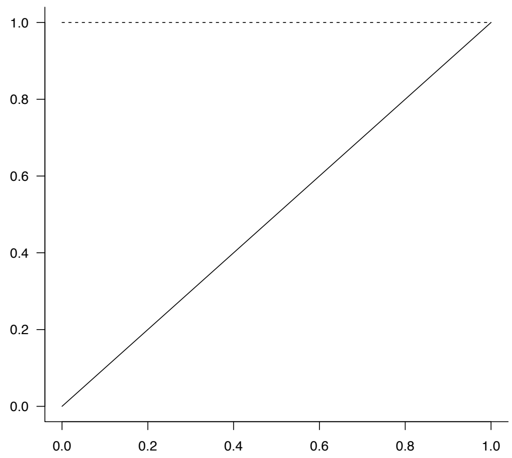


Figure 2.2: Basis of the linear model. Basis $B_1(x) = 1$ is symbolized by the dashed line and $B_2(x) = x$ by the solid line.

The simple linear model can be extended to numerous covariates (under the assumption that their influence can be approximated by an additive influence). If we suppose we have k covariates one has a $(n \times (k+1))$ dimensional design matrix $\boldsymbol{X}$ and correspondingly a $((k+1) \times 1)$ dimensional vector $\boldsymbol{\beta}$. To guarantee a unique solution of the minimization problem we need $\boldsymbol{X}$ to have full rank. Analogously, estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{y}}$ for the expanded model can be calculated from (2.3) and (2.4), respectively.

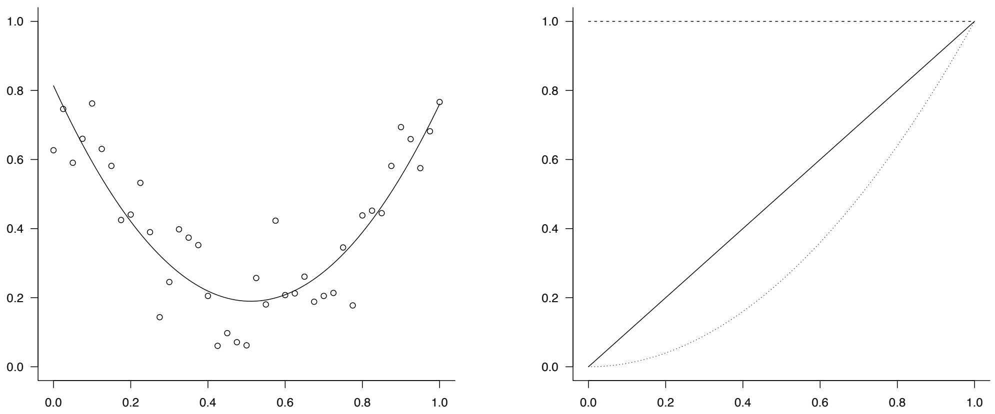
Looking again at (2.1) we notice that model (2.1) is a linear combination of 1 and x. This is why we can view 1 and x as basis functions

$$B_1(x) = 1 \qquad \text{and} \qquad B_2(x) = x$$

for model (2.1). Figure 2.2 shows the two basis functions.

Turning now to model and data in Figure 2.3(a) it becomes obvious that a linear model will not capture the features of the data adequately. We need other basis functions to obtain a good fit. To model the observable quadratic trend we introduce covariate x^2 which gets us to the quadratic model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i \tag{2.5}$$



(a) Quadratic model (b) Basis of the quadratic model

Figure 2.3: Quadratic model with corresponding basis. Basis $B_1(x) = 1$ is symbolized by the dashed line, $B_2(x) = x$ by the solid line and $B_3(x) = x^2$ by the dotted line.

with basis functions

$$B_1(x) = 1, \qquad B_2(x) = x \qquad \text{and} \qquad B_3(x) = x^2,$$

which can be seen in Figure 2.3(b).

The preliminary goal was to obtain a polynomial model. Both models (2.1) and (2.5) can be seen as a polynomial model: (2.1) is a polynomial model of degree one, the quadratic model (2.5) corresponds to a polynomial model of degree two. As a generalization we now define the polynomial model of degree p as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \epsilon_i. \qquad (2.6)$$

Analogously to the linear model we define the $p + 1$ basis functions

$$B_1(x) = 1, \qquad B_2(x) = x, \qquad \ldots, \qquad B_{p+1}(x) = x^p,$$

which characterize the polynomial model of degree p. The design matrix $\boldsymbol{X}$ is now given as

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{pmatrix}.$$

Estimations for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{y}}$ are obtainable through (2.3) and (2.4) with use of last-mentioned design matrix $\boldsymbol{X}$. Note that with the definitions of the basis functions we can rewrite model (2.6) as

$$y_i = f(x_i) = \beta_0 B_1(x_i) + \beta_1 B_2(x_i) + \cdots + \beta_p B_{p+1}(x_i) + \epsilon_i, \tag{2.7}$$

which will become useful in Section 2.2.1.

2.2 Penalized Spline Regression

2.2.1 B-splines

We will now introduce a type of basis functions which allows for a very flexible modeling of the dependency between y and x. Suppose we have K basis functions. As in equation (2.7) we write the model as

$$f(x) = \sum_{j=1}^{K} u_j B_j(x),$$

with u_j denoting the coefficient of basis function B_j. The definition of basis functions B_j (so-called B-spline basis functions) will be given in this subsection.

B-splines are described in detail by de Boor (2001), Eilers and Marx (1996), and Fahrmeir, Kneib, Lang, and Marx (2013). The principle idea is that one wants to obtain a smooth function by plugging pieces of a certain polynomial degree onto each other. For each polynomial of degree p, we want the resulting function f to be $(p-1)$-times continuous differentiable. Mathematically B-splines are defined recursively. The definition we give is the same as in Fahrmeir, Kneib, Lang, and Marx (2013) and is differing only in terms of notation from the definition of de Boor (2001). de Boor (2001) describes the numerical properties of B-splines and gives a detailed description of B-splines. Eilers and Marx (1996) encourage the use of B-splines in combination with an additional simple penalty in the context of regression and smoothing. They also give a good overview about B-splines.

A B-spline basis function of degree zero is defined as

$$B_j^0(x) = \mathbb{1}_{[\kappa_j,\kappa_{j+1})}(x) = \begin{cases} 1, & \text{if } \kappa_j \leq x < \kappa_{j+1}, \\ 0, & \text{otherwise,} \end{cases} \tag{2.8}$$

for $j = 1, \ldots, K-1$ and knots κ_j. A basis function of a general B-spline of degree $p \geq 1$ can now be defined as

$$B_j^p(x) = \frac{x - \kappa_{j-p}}{\kappa_j - \kappa_{j-p}} B_{j-1}^{p-1}(x) + \frac{\kappa_{j+1} - x}{\kappa_{j+1} - \kappa_{j+1-p}} B_j^{p-1}(x). \tag{2.9}$$

That means, each B-spline of degree p can be constructed from a B-spline of degree $p-1$ and each B-spline of arbitrary degree can be traced back to a B-spline of degree zero. In Figure 2.4 we plotted basis functions of various degrees. The figures were produced within the open-source software R (see R Core Team, R Core Team 2013) using function `splineDesign` (see also `?splineDesign` in R) of the R package `splines`. For construction of a single B-spline basis function of degree zero one needs two knots which are indicated by grey dotted lines. Looking at the basis function of degree one, we see that we need three knots for their construction: two outer knots and one inner knot. The next thing we notice is that the basis function of degree one is composed of two linear pieces joining each other in the inner knot. In general we can find that each single basis function of degree p consists of $p+1$ polynomial pieces of degree p which are defined by $p+2$ knots, from which p of them are inner knots. In Figures 2.4 and 2.5 all knots are chosen to be equidistant. As we mentioned earlier, we want a smooth construction which requires smooth connection of adjacent polynomial pieces. B-splines fulfil that condition as their first derivatives coincide in each knot (see de Boor, 2001). Also obvious from Figure 2.4 is that with increasing degree one gets a smoother behaviour.

Consider now more than one basis function as it is visualized in Figure 2.5. We see that each basis function overlaps with $2p$ basis functions as we count p overlapping basis function in each direction (except for basis functions near the boundary). A next thing we may see from Figure 2.5 is that for basis functions of degree p for a given value within the space spanned by the knots $p+1$ basis functions are positive. The grey solid lines in Figure 2.5 describe a further property of the B-spline basis functions as they visualize the sum of the basis functions. We see that for each value within the domain (indicated by dark dotted vertical lines) B-spline basis functions sum up to one. And we see that for a domain consisting of $\widetilde{K}$ knots, we need additionally $2p$ outer knots to construct a B-spline of degree p. To cover a domain parted into $\widetilde{K}$ knots, one needs $K = p + \widetilde{K} - 1$ basis functions of

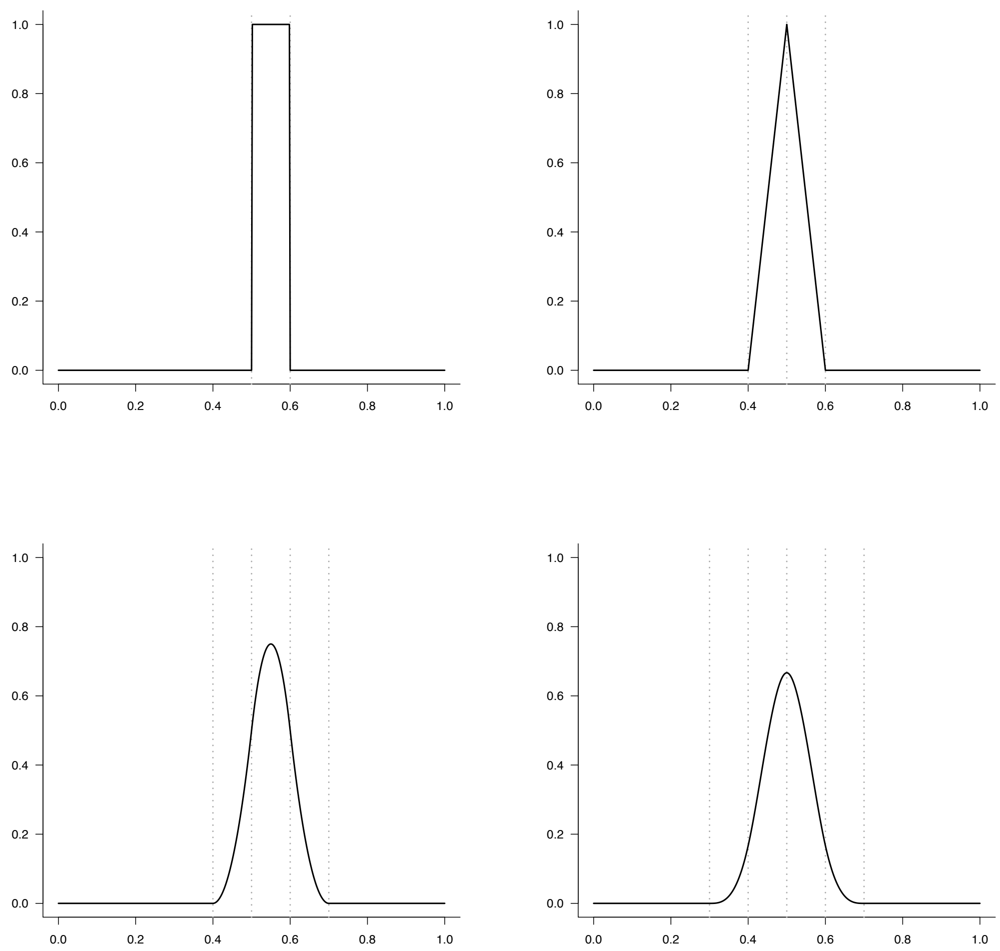


Figure 2.4: Single B-spline basis functions of degree $p = 0, 1, 2, 3$. The dotted lines correspond to knots needed for construction.

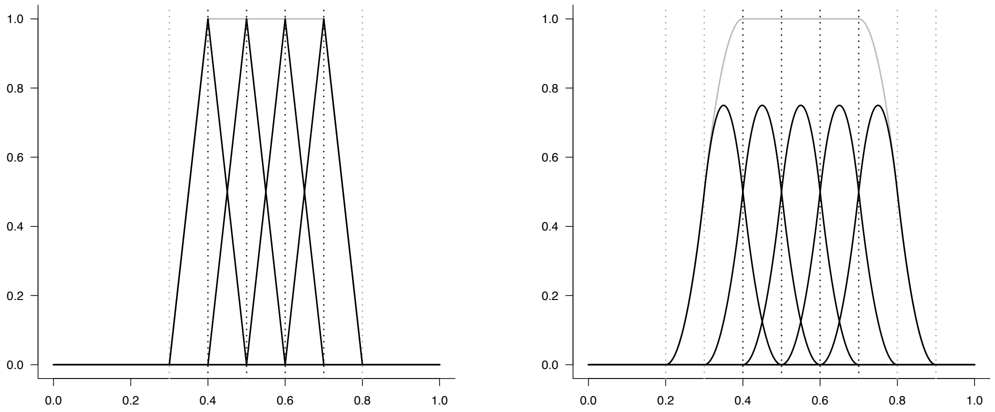


Figure 2.5: B-spline basis functions of degree $p = 1$ (left) and $p = 2$ (right) and their sum (grey solid line). Knots are visualized by the dotted lines. Lines for inner knots are drawn in black whereas for outer knots they are grey.

degree p.

B-splines have the nice feature, that the estimated coefficients of a B-spline not only give an estimation for function f, but also for the derivatives of f as we have

$$\frac{\partial f(x)}{\partial x} = p \sum_j \frac{\Delta u_j}{\kappa_j - \kappa_{j-l}} B_{j-1}^{p-1}(x), \tag{2.10}$$

where $\Delta u_j = u_j - u_{j-1}$ (see Fahrmeir, Kneib, Lang, and Marx, 2013). A proof is given in de Boor (2001). Equation (2.10) will become useful for defining a penalty matrix in Section 2.2.2. Constructing now the design matrix with the help of the basis functions leads to

$$\boldsymbol{B} := \begin{pmatrix} B_1(x_1) & \cdots & B_K(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \cdots & B_K(x_n) \end{pmatrix}.$$

To get an estimation for the coefficients u_j and to calculate fitted values $\hat{\boldsymbol{y}}$ we can still use equations (2.3) and (2.4) (by substitution of $\boldsymbol{X}$ through $\boldsymbol{B}$ and β_j through u_j).

2.2.2 P-splines

In Subsection 2.2.1 B-splines were defined, where the definition included the positioning of knots. A problem is, how many knots to choose and where to place them. The number of knots influences the flexibility of the fitted curve and the curve will behave more wiggly over parts of the domain where many knots are placed in contrast to parts where only few knots are placed. Eilers and Marx (1996) circumvent the problem by introducing a penalty to account for too much flexibility. This will be described in the following. The procedure will be to place a high number of knots equidistantly over the domain and to steer the smoothness by adding a penalization which accounts for too much deviation. Fahrmeir, Kneib, Lang, and Marx (2013) suggest to choose a number of knots between 20 and 40. Instead of minimizing $(\boldsymbol{y}-\boldsymbol{B}\boldsymbol{u})'(\boldsymbol{y}-\boldsymbol{B}\boldsymbol{u})$ itself, an additional penalty term is introduced. Let $\boldsymbol{K}$ be a symmetric penalty matrix. With

$$(\boldsymbol{y}-\boldsymbol{B}\boldsymbol{u})'(\boldsymbol{y}-\boldsymbol{B}\boldsymbol{u})+\lambda\boldsymbol{u}'\boldsymbol{K}\boldsymbol{u} \tag{2.11}$$

where $\lambda \geq 0$ we get a penalized least squares criterion which is minimized with respect to $\boldsymbol{u}$. The amount of smoothness is now controlled by λ. The problem of determining smoothing parameter λ is discussed in Section 2.2.3. Differentiating (2.11) and solving the normal equations yields

$$\hat{\boldsymbol{u}} = (\boldsymbol{B}'\boldsymbol{B}+\lambda\boldsymbol{K})^{-1}\boldsymbol{B}'\boldsymbol{y}, \tag{2.12}$$

which is equal to equation (2.3) for $\lambda = 0$. For $\lambda \neq 0$ both estimators differ through the additional term $\lambda\boldsymbol{K}$. Fitted values $\hat{\boldsymbol{y}}$ are obtained by calculating

$$\hat{\boldsymbol{y}} = \boldsymbol{B}\hat{\boldsymbol{u}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B}+\lambda\boldsymbol{K})^{-1}\boldsymbol{B}'\boldsymbol{y}. \tag{2.13}$$

The penalty matrix $\boldsymbol{K}$ is constructed such that a strong variation in the estimator $\hat{f}$ is punished. O'Sullivan (1986, 1988) proposed to use the second derivative in its capacity as a measure for the curvature of a function as penalty term. This leads to a penalty of the form

$$\lambda \int (f''(z))^2\, dz. \tag{2.14}$$

Eilers and Marx (1996) use equation (2.10) to construct a simple, but effective penalty matrix based on differences of adjacent coefficients of B-splines. They show that there exists a strong connection between a penalization of second order differences and a penalization of the second derivative. Therefore, considering a penalty term based on second order differences leads to

$$\lambda \sum_{j=3}^{K} (\Delta^2 u_j)^2, \tag{2.15}$$

where $\Delta^2 u_j = u_j - 2u_{j-1} + u_{j-2}$. We are now able to construct from (2.15) the penalty matrix $\boldsymbol{K}$ with the help of difference matrix $\boldsymbol{D}$ with dimensions $(K-2) \times K$, where

$$\boldsymbol{D} = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Setting now $\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{D}$ leads to

$$\lambda \sum_{j=3}^{K} (\Delta^2 u_j)^2 = \lambda \boldsymbol{u}'\boldsymbol{K}\boldsymbol{u}.$$

which gives the desired penalty (2.15). In the next subsection the selection of λ is discussed briefly.

2.2.3 Reparameterization of P-splines

One way to choose the smoothing parameter λ is to reformulate the penalized spline model as a mixed model. A simple mixed model is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}) \text{ and } \boldsymbol{\gamma} \sim \mathrm{N}(\mathbf{0}, \sigma_\gamma^2 \boldsymbol{I}), \tag{2.16}$$

where matrix $\boldsymbol{U}$ is constructed similarly to $\boldsymbol{X}$. In a mixed model we have a parameter vector $\boldsymbol{\gamma}$, which is not regarded as fixed, but random. This becomes useful when, for example, one looks at data with repeated measurements. The mixed model allows to introduce correlations between subjects which is obviously necessary as measurements from the same subject can not be treated as independent observations. A random intercept, for example, allows for a shift for each individual. In this case we have a 1 in cell (j, i) of matrix $\boldsymbol{U}$ for the j-th observation of subject i and 0 otherwise (with $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$). A description of mixed models can be found in Verbeke and Molenberghs (2000) and Diggle, Liang, and Zeger (1994). The link between P-spline regression and mixed models is described in Ruppert, Wand, and Carroll (2003) and Fahrmeir, Kneib, Lang, and Marx (2013).

For mixed models one obtains

$$(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{U\gamma})'(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{U\gamma}) + \frac{\sigma_\epsilon^2}{\sigma_\gamma^2}\boldsymbol{\gamma}'\boldsymbol{\gamma} \tag{2.17}$$

as penalized least squares criterion in contrast to equation (2.11) for penalized splines. By setting $\boldsymbol{Z} := [\boldsymbol{X}\ \boldsymbol{U}]$ and $\boldsymbol{v} := [\boldsymbol{\beta}\ \boldsymbol{\gamma}]$ one gets for (2.17)

$$(\boldsymbol{y} - \boldsymbol{Zv})'(\boldsymbol{y} - \boldsymbol{Zv}) + \frac{\sigma_\epsilon^2}{\sigma_\gamma^2}\boldsymbol{\gamma}'\boldsymbol{\gamma}. \tag{2.18}$$

Remembering equation (2.11) the goal now is to obtain a partition of $\boldsymbol{u}$ into a fixed part $\boldsymbol{u}_{\text{fix}}$ and a random part $\boldsymbol{u}_{\text{random}}$. Following Fahrmeir, Kneib, Lang, and Marx (2013) we set

$$\boldsymbol{u} = \tilde{\boldsymbol{X}}\boldsymbol{u}_{\text{fix}} + \tilde{\boldsymbol{U}}\boldsymbol{u}_{\text{random}}$$

and choose $\tilde{\boldsymbol{X}}$ such that $\tilde{\boldsymbol{X}}$ and $\boldsymbol{K}$ are orthogonal and $\tilde{\boldsymbol{U}}$ such that $\tilde{\boldsymbol{U}}'\boldsymbol{K}\tilde{\boldsymbol{U}} = \boldsymbol{I}$. Note that the number of rows for both $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{U}}$ is equal to the number of elements of $\boldsymbol{u}$ which is K whereas the number of columns depends on the rank of the penalty matrix $\boldsymbol{K}$. With second order differences we have rank$(\boldsymbol{K}) = K - 2$ and therefore we obtain $K \times 2$ as dimension for $\tilde{\boldsymbol{X}}$ and $K \times (K-2)$ as dimension for $\tilde{\boldsymbol{U}}$. A valid choice for $\tilde{\boldsymbol{U}}$ for example is to set $\tilde{\boldsymbol{U}} = \boldsymbol{D}'(\boldsymbol{DD}')^{-1}$. Considering now the penalty term $\boldsymbol{u}'\boldsymbol{Ku}$ we get

$$\boldsymbol{u}'\boldsymbol{Ku} = \boldsymbol{u}_{\text{random}}'\boldsymbol{u}_{\text{random}}$$

and for the model

$$\boldsymbol{y} = \boldsymbol{Bu} + \boldsymbol{\epsilon} = \boldsymbol{B}(\tilde{\boldsymbol{X}}\boldsymbol{u}_{\text{fix}} + \tilde{\boldsymbol{U}}\boldsymbol{u}_{\text{random}}) + \boldsymbol{\epsilon} = \check{\boldsymbol{X}}\boldsymbol{u}_{\text{fix}} + \check{\boldsymbol{U}}\boldsymbol{u}_{\text{random}} + \boldsymbol{\epsilon}$$

with $\check{\boldsymbol{X}} := \boldsymbol{B}\tilde{\boldsymbol{X}}$ and $\check{\boldsymbol{U}} := \boldsymbol{B}\tilde{\boldsymbol{U}}$. Smoothing parameter λ now can be determined using equation (2.18) by setting $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\gamma^2}$. This will be used in Section 4 where we describe how to estimate σ_ϵ^2 and σ_γ^2 using the Schall algorithm as proposed by Schall (1991) and for expectile regression by Schnabel and Eilers (2009b).

2.3 Semiparametric Regression

So far, we mostly considered a single explanatory variable. The extension to numerous explanatory variables is straightforward. The covariates with a linear effect on response y

are denoted as $\boldsymbol{x}_{\text{para}} = (x_1, \dots, x_{\tilde{d}})'$ and are gathered in design matrix $\boldsymbol{X}$ with

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1\tilde{d}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{n\tilde{d}} \end{pmatrix}.$$

The nonlinear effects $\boldsymbol{x}_{\text{nonpara}} = (x_{\tilde{d}+1}, \dots, x_d)'$ are gathered in design matrix $\boldsymbol{B}$ with $\boldsymbol{B}$ now defined as $\boldsymbol{B} = (\boldsymbol{B}_{\tilde{d}+1}, \dots, \boldsymbol{B}_d)$ where each $\boldsymbol{B}_j$ can be obtained as described in Subsection 2.2.1 and reparameterized as in Subsection 2.2.3. Setting $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{B}_{\tilde{d}+1}, \dots \boldsymbol{B}_d)$ allows to get an estimator $\hat{\boldsymbol{v}}$ as

$$\hat{\boldsymbol{v}} = (\boldsymbol{Z}'\boldsymbol{Z} + \boldsymbol{K})^{-1} \boldsymbol{B}'\boldsymbol{y}.$$

Coefficient vector $\boldsymbol{v}$ consists of a parametric part $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{\tilde{d}})$ and a nonparametric part $\boldsymbol{u}_{\tilde{d}+1}, \dots, \boldsymbol{u}_d$. Penalty matrix $\boldsymbol{K}$ needs to be redefined as

$$\boldsymbol{K} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \lambda_{\tilde{d}+1}\boldsymbol{K}_{\tilde{d}+1} & & \\ & & & & \ddots & \\ & & & & & \lambda_d \boldsymbol{K}_d \end{pmatrix}$$

where the blank space needs to be filled with zeros. Each penalty term $\lambda_j \boldsymbol{K}_j$ is constructed as in Subsection 2.2.2. The term semiparametric indicates that we have a model which contains both parametric and nonparametric parts.

2.4 Expectiles and Expectile Regression

In the previous section an introduction into ordinary least squares regression was given where the mean of the response, given covariates, was modeled. Another, more general approach is to not only model one single curve, but a variety of curves to additionally describe non-central parts of the data. This is done in expectile and quantile regression. Note that both expectile and quantile estimation are special cases of so called M-quantile estimation as described in Breckling and Chambers (1988).

Koenker and Bassett (1978) develop the concept of quantile regression. An overview is given by Koenker (2005). In quantile regression one models

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

where $\alpha \in (0,1)$ and $q_{i,\alpha}$ is the α-quantile. We need that $P(\epsilon_{i,\alpha} \leq 0) = \alpha$ holds i.e. that the α-quantile of $\epsilon_{i,\alpha}$ is zero. The α-quantile may be modeled in dependency of covariates x_i (in analogy to the previous sections). Defining asymmetric weights $w_{i,\alpha}$ as

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

allows to calculate estimates for $q_{i,\alpha}$ by minimizing the weighted sum of absolute differences

$$\sum w_{i,\alpha} \left| y_i - q_{i,\alpha} \right| . \tag{2.19}$$

As the expression in (2.19) is not differentiable, linear programming is used to obtain quantile estimates. A description can be found in Koenker (2005).

Aigner, Amemiya, and Poirier (1976) and Newey and Powell (1987) replace the L_1 distance by a L_2 distance which leads to a minimization problem which is differentiable and can therefore easily be solved. An α-expectile is defined by minimization of

$$\sum w_{i,\alpha} \left(y_i - m_{i,\alpha} \right)^2 \tag{2.20}$$

with respect to $m_{i,\alpha}$. Weights $w_{i,\alpha}$ are as defined above and $\alpha \in (0,1)$. Again, $m_{i,\alpha}$ may depend on values x_i. Expectile regression regained some interest in the last years as can be seen by the increasing number of publications on expectiles like Schnabel and Eilers (2009b), De Rossi and Harvey (2009), Sobotka and Kneib (2012), Guo and Härdle (2013) and Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013). As stated above, equation (2.20) can be minimized by differentiation but depends on weights $w_{i,\alpha}$ which again depend on $m_{i,\alpha}$. The remedy of this problem is to start with equal weights of 0.5 for α to get a first estimation of $\hat{m}_{i,\alpha}$. Afterwards the weights $w_{i,\alpha}$ can be calculated to allow for an update of $\hat{m}_{i,\alpha}$. This procedure is iterated until convergence of the weights.

Often a so called check function is used to define quantiles (see e.g. Koenker and Bassett, 1978). The weighted sum of absolute differences can be expressed as

$$\sum \rho_\alpha(y_i - q_{i,\alpha})$$

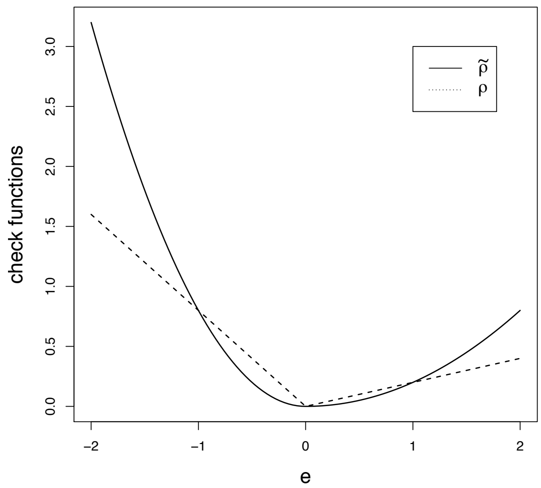


Figure 2.6: Check function $\tilde{\rho}$ for expectiles and ρ for quantiles (for $\alpha = 0.2$)

with check function ρ_α defined by

$$\rho_\alpha(e) = \begin{cases} e\alpha, & \text{for } e \geq 0, \\ e(\alpha - 1), & \text{for } e < 0. \end{cases}$$

For expectile estimation we need to exchange check function ρ with $\tilde{\rho}$ where $\tilde{\rho}$ is given by

$$\tilde{\rho}_\alpha(e) = \begin{cases} e^2\alpha, & \text{for } e \geq 0, \\ e^2(1 - \alpha), & \text{for } e < 0. \end{cases}$$

A plot of both functions can be seen in Figure 2.6. From Figure 2.6 we can also deduce some properties of quantile and expectile regression. We can see, why the term in equation (2.19) is not differentiable as we see a peak at $e = 0$. Also, one can see that for values larger than one, $\tilde{\rho}$ takes higher values than ρ where the difference increases with e. That shows that expectiles are more influenced by outliers than quantiles.

Criticism passed on expectiles often concerns the lack of interpretability. There are several remedies to circumvent this problem. First, there is the suggestion to estimate not only one expectile curve but a whole set of them and to interpret the set of expectile curves. Parallel expectile curves for example imply homoscedasticity. Also skewness can be detected comparing the differences between neighboring lower expectile curves and neighboring higher expectile curves. In principle the whole distribution of the response (conditional on x) is

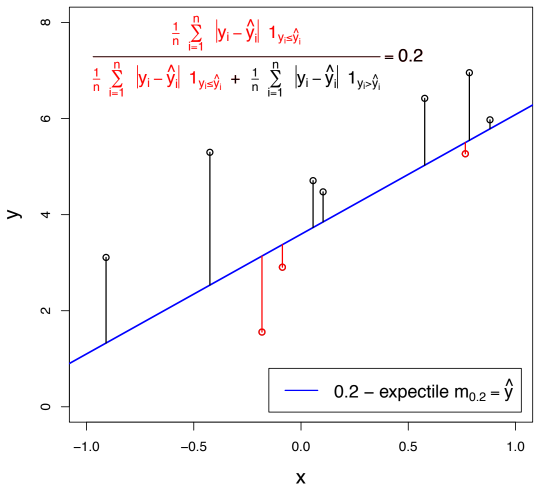


Figure 2.7: Interpretation of 0.2–expectile $m_{0.2}$.

described by expectiles. Second, one can transform expectiles into quantiles without loosing efficiency. This is shown in Section 3 where the method is described and applied. Third, there is a sensible interpretation of expectiles as we will see in the following. The interpretation of quantiles is quite easy. For the α-quantile q_α we have, that a proportion of $100\alpha\%$ of the data lies below q_α and a proportion of $100(1-\alpha)\%$ of the data lies above q_α. At first glance the interpretation of expectiles is not that intuitive. Nevertheless, we try to visualize the meaning of the 0.2-expectile in Figure 2.7. Yao and Tong (1996) state, that, for given x, the α-expectile m_α is determined such that $100\alpha\%$ of the mean distance between m_α and y is given by the mass below it and correspondingly $100(1-\alpha)\%$ of the mean distance between m_α and y is given by the mass above it. This exact property can be seen in Figure 2.7.

Chapter 3

Expectile and Quantile Regression - David and Goliath?

This chapter is developed in joint work with Fabian Sobotka, Thomas Kneib and Göran Kauermann. The final, definitive version of this paper will be published in Statistical Modelling by SAGE Publications India Pvt Ltd, Linda Schulze Waltrup and Göran Kauermann developed jointly the connection between expectiles and quantiles and the monotone expectile sheets. They proposed a method to estimate quantiles from a set of expectiles. Linda Schulze Waltrup implemented the method and ran an extensive empirical evaluation between quantiles and expectiles and the recurrent example of the rent index of Munich. Göran Kauermann investigated the behaviour of expectiles and quantiles in the tail. Fabian Sobotka ran the simulation study where crossings between neighboring quantile and expectile curves were explored. Also, the example of the expected shortfall was constructed by Fabian Sobotka. All authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript.

Abstract

Recent interest in modern regression modelling has focused on extending available (mean) regression models by describing more general properties of the response distribution. A completely distribution free 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 paper. 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 neighboring curves may cross each other. We propose a 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 paper.

3.1 Introduction

Quantile Regression allows to estimate 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, $i = 1, \dots, n$ 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

$$\sum_{i=1}^{n} w_{i,\alpha} \left| y_i - q_{i,\alpha} \right| \tag{3.1}$$

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, (3.1) can be minimized by linear programming, see e.g. Koenker (2005).

As an alternative to quantile regression, Aigner, Amemiya, and Poirier (1976) and Newey and Powell (1987) proposed to replace the L_1 distance in (3.1) by a quadratic L_2 term leading to the asymmetric least squares

$$\sum_{i=1}^{n} w_{i,\alpha} \left(y_i - m_{i,\alpha} \right)^2 \tag{3.2}$$

where the minimizer $\hat{m}_{i,\alpha}$, say, is called (estimated) expectile. 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, Ranalli, and Salvati, 2009; Sobotka and Kneib, 2012; Guo and Härdle, 2013). An overview about methods focussing on estimation procedures regarding more features of the data than its center (including semiparametric expectile and quantile regression) can be found in Kneib (2013).

In this paper, 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 to 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 (3.1) to (3.2), 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 $h : (0, 1) \to (0, 1)$ such that $q_\alpha = m_{h(\alpha)}$, where $h(.)$ is defined through

$$h(\alpha) = \frac{-\alpha q_\alpha + G(q_\alpha)}{-m_{0.5} + 2G(q_\alpha) + (1 - 2\alpha) q_\alpha} \tag{3.3}$$

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 paper we will make use of relation (3.3) 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 (3.3). One of the key findings of the paper is that estimates $\hat{m}_{\hat{h}(\alpha)}$ 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 so called crossing quantile functions. These occur if for estimated quantiles one gets $\hat{q}_{\alpha_1}(x) > \hat{q}_{\alpha_2}(x)$ for $\alpha_1 < \alpha_2$ for some value x (in the observed range of the covariate), where $\hat{q}_\alpha(x) = \hat{\beta}^{(q)}_{0\alpha} + x\hat{\beta}^{(q)}_{1\alpha}$. Several methods, algorithms and model constraints have been proposed to circumvent the problem. Bondell, Reich, and Wang (2010) make use of linear programming. They also give a good overview about earlier proposals including Koenker (1984), He (1997), Wu and Liu (2009) or Neocleous and Portnoy (2008). Chernozhukov, Fernández-Val, and Galichon (2010) rearrange the fitted (linear) curves into a set of non-crossing curves. Schnabel and Eilers (2013b) propose 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 of expectiles occurs 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, Ng, and Portnoy (1994) proposed spline based estimation. Bollaerts, Eilers, and Aerts (2006) make use of penalized B-splines with an L_1 penalty. Recently Reiss and Huang (2012) suggest quantile estimation based on penalized iterative least squares (see also Yuan, 2006). For expectiles, smooth estimation has been pursued by e.g. Pratesi, Ranalli, and Salvati (2009), and Schnabel and Eilers (2009b). 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 to incorporate non-crossing conditions, as in Bondell, Reich, and Wang (2010) for quantile estimation. Schnabel and Eilers (2013b) propose quantile sheets based on penalized spline smoothing, see also Schnabel and Eilers, 2014. We extend and modify these ideas.

Expectiles might not gain the popularity as quantiles, but we think they deserve their niche. For example, Aigner, Amemiya, and Poirier (1976) constructed 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 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 is a trimmed mean, that is the mean of a random variable conditional that its value is above (or below) a certain quantile. The expected shortfall can be written as a function of both, the quantile and the expectile for a level α. Estimation of the expected shortfall has been recently proposed by Leorato, Peracchi, and Tanase (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 expected shortfall. This connection becomes extremely useful for calculating the expected shortfall as it both depends on expectiles and their corresponding quantiles (as described by Taylor, 2008).

The paper is organized as follows. In Section 3.2 we compare and contrast quantiles and expectiles, both theoretically and simulation based. In Section 3.3 we look at quantile and expectile regression before Section 3.4 provides extensions and examples. Section 3.5 concludes the contest of David and Goliath.

3.2 Expectiles and Quantiles

3.2.1 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 (3.3) can be derived numerically, which in practice allows to calculate the quantile function from a fitted expectile function $\hat{m}(\alpha)$, say. Note first that expectiles are defined through

$$m_\alpha = \frac{(1-\alpha)G(m_\alpha) + \alpha(m_{0.5} - G(m_\alpha))}{(1-\alpha)F(m_\alpha) + \alpha(1 - F(m_\alpha))} \tag{3.4}$$

which needs to be solved numerically for $F(m_\alpha)$. Let therefore $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_L < 1$ 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 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 3.2.2 and in the examples in Sections 3.2.4 and 3.4.2 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 one) then there is obviously just a small difference between the minimal expectile and the minimal observed value of the data.

We now solve (3.4) 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

$$\hat{F}_l := \hat{F}_m(\hat{m}_l) = \sum_{j=1}^{l} \hat{\zeta}_j \tag{3.5}$$

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

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

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 $\hat{\boldsymbol{\zeta}} = (\hat{\zeta}_1, \ldots, \hat{\zeta}_L)$ from a set of estimated expectiles. Details are given in the Appendix.

Defining the linear interpolation $\hat{F}_m(y) = \sum_{j=1}^{l} \hat{\zeta}_j + \hat{\zeta}_{j+1}(y - \hat{m}_l)/(\hat{m}_{l+1} - \hat{m}_l)$ for $y \in [\hat{m}_l, \hat{m}_{l+1})$ allows 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 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 $h(.)$.

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

$$\lambda_{\text{pen}} \sum_{l=1}^{L-1} \left(\frac{\hat{\zeta}_l}{\hat{m}_l - \hat{m}_{l-1}} - \frac{\hat{\zeta}_{l+1}}{\hat{m}_{l+1} - \hat{m}_l} \right)^2 . \tag{3.6}$$

Details are provided in the Appendix. Note that the calculation of quantiles from expectiles is somewhat numerically demanding. As alternative there is also a more naïve 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 expected shortfall. 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 paper later.

3.2.2 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 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 example for $n = 499$, 459 expectiles were estimated). We simulated (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 part of the R package `expectreg` (and 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 (3.6) does not only lead to a smooth distribution function and therefore to smooth quantile estimates, but also improves the numerical stability of the calculations.

In Figure 3.1 we show exemplary for one 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 3.4.3 where we estimate the expected shortfall.

In Figure 3.2 we visualise 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

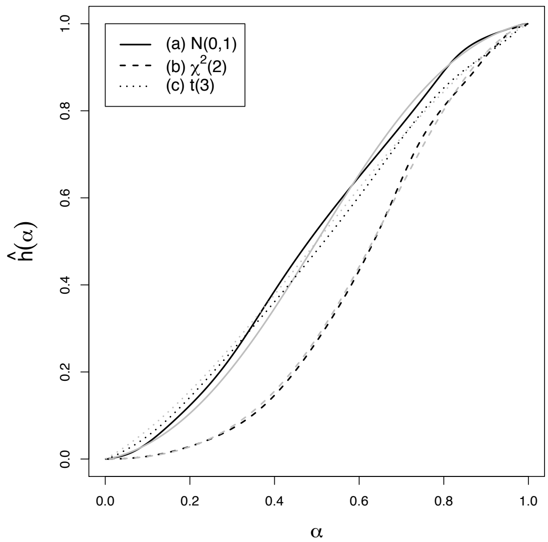


Figure 3.1: Estimated transfer function $\hat{h}(.)$ (in black) and theoretical transfer function (in grey) for the three kind of error distributions considered here.

$n = 199$ and $n = 499$ are very similar, Figure 3.2 concentrates on $n = 199$. The first plot of Figure 3.2 shows the relative RMSE (root mean squared error) 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 3.2 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 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 we do not report here as the results are quite similar.

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

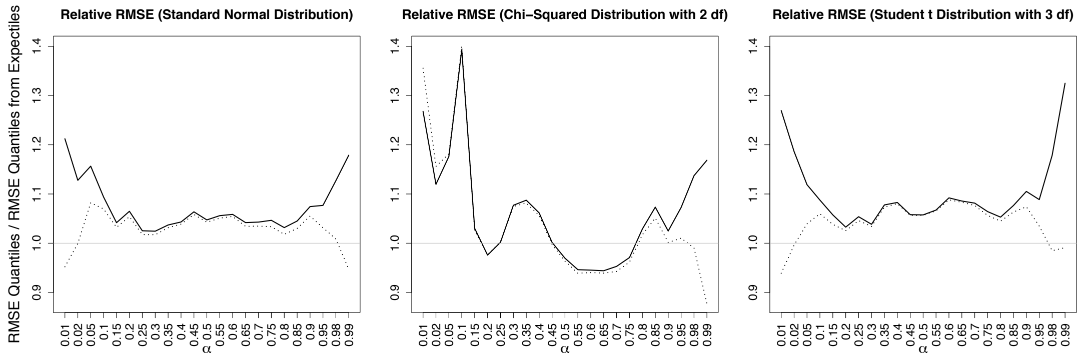


Figure 3.2: Relative Root Mean Squared Error $RMSE(\hat{q}_\alpha)/RMSE(\hat{m}_{\hat{h}(\alpha)})$ (solid lines) and $RMSE(\hat{q}_\alpha^{\text{smooth}})/RMSE(\hat{m}_{\hat{h}(\alpha)})$ (dotted lines) where $h(.)$ is estimated through $\hat{h}(.)$ for different simulation distributions and $n = 199$.

3.2.3 Expectiles and Quantiles in the tail

As can be seen from Figure 3.1 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

$$\alpha = \xi/n \text{ (or } \alpha = 1 - \xi/n) \tag{3.7}$$

for some $\xi \geq 1$. 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 minimizer of (3.2) for $\tilde{\alpha} = h(\alpha)$ and we get

$$\hat{m}_{\tilde{\alpha}} = \Big(\sum_{i=1}^{n} \hat{w}_{i,\tilde{\alpha}}\Big)^{-1} \Big(\sum_{i=1}^{n} \hat{w}_{i,\tilde{\alpha}} Y_i\Big) \tag{3.8}$$

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 (3.8) is not an analytic definition, since the iterated weights $\hat{w}_i$ depend on the fitted value $\hat{m}_{\tilde{\alpha}}$. We simplify (3.8) 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 to approximate (3.8) to

$$\hat{m}_{\tilde{\alpha}} \approx \Big(\sum_{i=1}^{n} w_{i,\tilde{\alpha}}\Big)^{-1}\Big(\sum_{i=1}^{n} w_{i,\alpha} Y_i\Big). \tag{3.9}$$

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 (3.7) we may approximate (3.9) through expansion to

$$\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}} \approx \xi^{-1} \sum_{i=1}^{n} w_{i,\tilde{\alpha}}(Y_i - m_{\tilde{\alpha}}) - \tag{3.10}$$

$$\xi^{-2} \sum_{i=1}^{n} (w_{i,\tilde{\alpha}} - w) \sum_{j=1}^{n} w_j (Y_j - m_{\tilde{\alpha}}) + \ldots \quad . \tag{3.11}$$

The first component in (3.10) has mean zero and variance

$$V_{\tilde{\alpha}} := \xi^{-2}\Big[(1-\tilde{\alpha})^2\Big\{H(q_\alpha) - 2G(q_\alpha)q_\alpha + q_\alpha^2\alpha\Big\} \tag{3.12}$$

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

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 (3.11) we find that its mean equals $\xi^{-2}\tilde{\alpha}(q_\alpha - m_{0.5})$ while its variance is of order $O(\tilde{\alpha}^2)O(V_\alpha)$. Arguing that $\tilde{\alpha} \ll \alpha$, see Appendix, we can conclude that the second term in (3.11) is of ignorable asymptotic order for $\alpha = \xi/n \to 0$. The same holds by simple calculation for the remaining components not explicitly listed in (3.10) and (3.11). Hence, we may approximate the distribution of $\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ through

$$\hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}} \approx \xi^{-1} \sum_{i=1}^{n} w_{i,\tilde{\alpha}}(y_i - m_{\tilde{\alpha}}). \tag{3.13}$$

In particular, with (3.13), we get the (asymptotic) unbiasedness $E(\hat{m}_{\tilde{\alpha}}) = m_{\tilde{\alpha}}$. One may even derive asymptotic normality from (3.13) 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

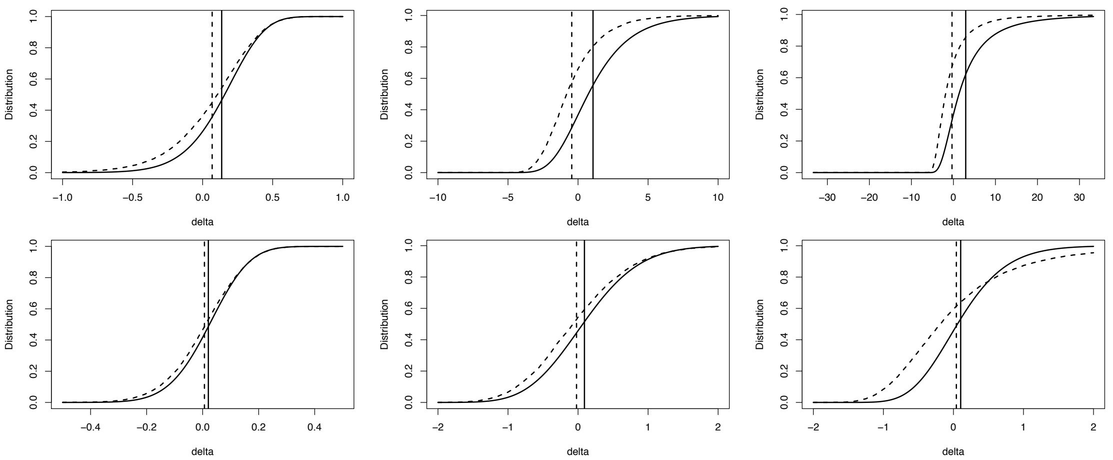


Figure 3.3: Sampling distribution of $\delta = \hat{q}_\alpha - q_\alpha$ (solid line) and $\delta = \hat{m}_{\tilde{\alpha}} - m_{\tilde{\alpha}}$ (dashed line) for different underlying distribution functions. Top row is for $\alpha = 0.999$ and bottom row for $\alpha = 0.99$. Left column is for normal distribution, middle column for Chi-squared (df=2) distribution and right column for t-distribution (df=3). Vertical line indicates the mean values for both distributions.

quantile estimation in the tails of the distribution. Following Koenker and Bassett (1978) and Koenker (2005, page 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 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

$$P(\hat{q}_\alpha - q_\alpha \leq \delta) = P\big(\sum_{i=1}^n 1\{Y_i \leq q_\alpha + \delta\} \geq n\alpha\big) = 1 - P(Z_\delta < n\alpha) \tag{3.14}$$

where Z_δ is a binomial random variable with parameters $Z_\delta \sim Bin\big(n, F(q_\alpha + \delta)\big)$. Note that $F(q_\alpha + \delta) \approx \alpha + f(q_\alpha)\delta$ and with (3.7) 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 $1 - P(Z \leq \xi)$ for $Z \sim Poisson(\xi)$. For instance for $\xi = 1$ 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 $\hat{q}_\alpha$.

We run a small simulation to study the performance of tail expectile and tail quantile estimation, respectively. We simulate data and look at the distribution of extreme quantiles. In Figure 3.3 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) quantile and the corresponding expectile. We simulate from (a) a normal distribution

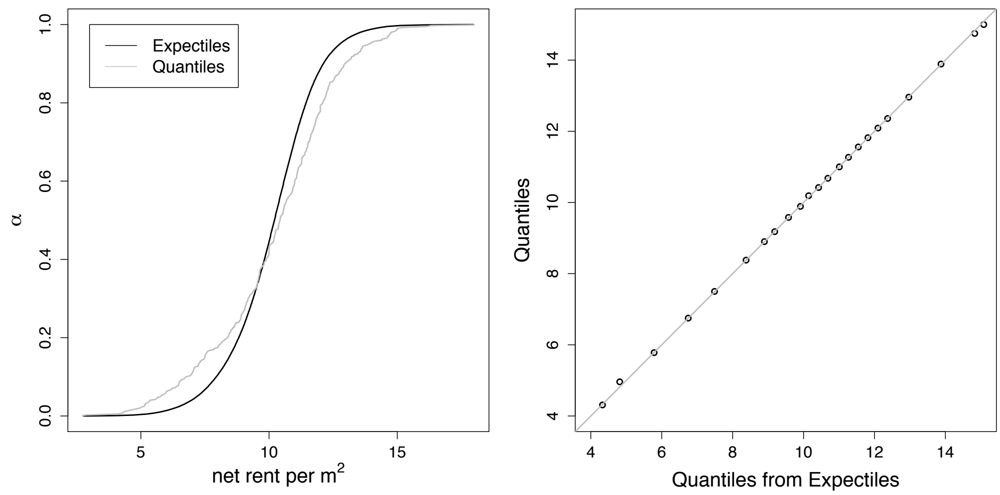


Figure 3.4: Left: Quantiles and expectiles for the Munich rent data. Right: QQ-Plot quantiles vs. quantiles from expectiles.

(left column), (b) a Chi-squared (with 2 degrees of freedom, middle column) and (c) a t-distribution (with 3 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 one (or close to zero).

Conclusion: Overall we may conclude that expectile 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.

3.2.4 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 analyse the variable giving the net rent per square meter (m^2) for each apartment. For illustration we restrict our attention to apartments between 45 m^2 and 55 m^2 and examine the net rent per square meter 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 3.4 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 right part of Figure 3.4). 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 the fluctuation around the identity line becomes larger. This behaviour is also supported by our simulation results in Section 3.2.2.

3.3 Quantile and Expectile Regression

3.3.1 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, respectively. 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 (3.1) and the corresponding L_2 version (3.2), 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 are crossing fitted functions. For $0 < \alpha_1 < \alpha_2 < 1$, by definition, we have $q_{\alpha_1}(x) < q_{\alpha_2}(x)$ and $m_{\alpha_1}(x) < m_{\alpha_2}(x)$, respectively, 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 and some of them 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 neighboring fitted expectiles and quantiles, respectively. 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

$$y = 4 + 3x + \epsilon. \tag{3.15}$$

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), respectively. 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 analysed 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 equation (3.3), which makes the estimates comparable. For each data set and every generated covariate value, all neighboring pairs of α and $h(\alpha)$ are checked for crossing regression lines within the range of observed covariates.

expectiles

$\epsilon \sim$	$N(0,1.5^2)$			$\chi^2(2)$			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

$\epsilon \sim$	$N(0,1.5^2)$			$\chi^2(2)$			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

Table 3.1: Number of crossings between two neighboring expectiles / quantiles in the linear model (3.15) from 1000 data sets, starting with $\alpha = 0.01$. Crossing counts are given for all ten 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.

The resulting number of crossings within the 1000 replications is summarized in Table 3.1. 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 quantile.

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

3.3.2 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, Reich, and Wang (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 $q_\alpha(x)$ and $m_\alpha(x)$ with bivariate functions

$$q(\alpha, x) \text{ and } m(\alpha, x) \tag{3.16}$$

where both functions are smooth (or just linear) in direction of α and smooth (or just linear) in direction of x. The bivariate functions may be called *quantile sheets* or *expectile sheets*. The setting (3.16) transfers the estimation exercise to bivariate smoothing, as proposed in Schnabel and Eilers (2013b). We replace (or approximate) $q(\alpha, x)$ through

$$q(\alpha, x) = [\boldsymbol{B}^{(1)}(\alpha) \otimes \boldsymbol{B}^{(2)}(x)]\ \boldsymbol{a} \tag{3.17}$$

where $\boldsymbol{B}^{(1)}(\alpha)$ is a (linear) B-spline basis set up on knots $0 < \alpha_1 < \ldots < \alpha_L < 1$ and $\boldsymbol{B}^{(2)}(x)$ is a B-spline basis built upon some knots $\kappa_1 < \ldots < \kappa_K$ covering the range of observed values of x and $\boldsymbol{a}$ 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 $\boldsymbol{a}$ may then be indexed by a_{lk} for $l = 1, \ldots, L$, $k = 1, \ldots, K$ and let $\boldsymbol{a}_{l.} = (a_{l1}, \ldots, a_{lk})'$. Non-crossing quantiles are now guaranteed by linear constraints on the parameter vector of the form

$$\boldsymbol{B}^{(2)}(x)(\boldsymbol{a}_{l.} - \boldsymbol{a}_{l+1.}) \leq 0 \text{ for } l = 1, \ldots, L-1 \tag{3.18}$$

for all x in the observed range of the covariates. If $\boldsymbol{B}^{(2)}(x)$ is a linear B-spline basis this simplifies to $a_{lk} \leq a_{l+1k}$ for $l = 1, \ldots, L-1$ and $k = 1, \ldots, K$. In general we can formulate (3.18) as linear constraint by inserting for x the observed values $x_i, i = 1, \ldots, n$.

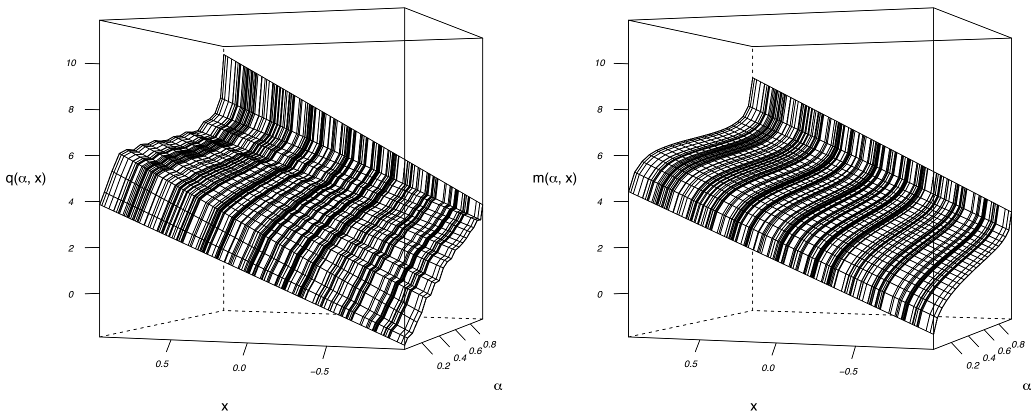


Figure 3.5: Quantile and expectile sheets for normally distributed errors.

We can now fit $q(\alpha, x)$ by replacing (3.1) with its multiple version

$$\sum_{l=1}^{L}\sum_{i=1}^{n} w_{i,\alpha_l}\left|y_i - \boldsymbol{B}^{(1)}(\alpha_l) \otimes \boldsymbol{B}^{(2)}(x_i)\ \boldsymbol{a}\right| \tag{3.19}$$

which is minimized with respect to $\boldsymbol{a}$ subject to the linear constraints (3.18) 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 function $q(\alpha, x)$.

Replacing the L_1 distance in (3.19) by the L_2 distance

$$\sum_{l=1}^{L}\sum_{i=1}^{n} w_{i,\alpha_l}\Big(y_i - \boldsymbol{B}^{(1)}(\alpha_l) \otimes \boldsymbol{B}^{(2)}(x_i)\ \boldsymbol{u}\Big)^2 \tag{3.20}$$

gives a weighted least squares criterion which allows to estimates the expectile sheet $m(\alpha, x) = \boldsymbol{B}^{(1)}(\alpha) \otimes \boldsymbol{B}^{(2)}(x)\ \boldsymbol{u}$, where again the linear constraints (3.18) need to be fulfilled. Estimation can be carried out by iterative quadratic programming. See also Schnabel and Eilers (2013a) for a (more restrictive) approach to obtain non-crossing expectile curves.

In Figure 3.5 we show exemplary for the simulation model (3.15) of the previous subsection the resulting quantile and expectile sheets for 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 rough 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 3.2 to derive quantiles based on expectiles. The code for calculating the linear non-crossing quantile curves by Bondell, Reich, and Wang (2010) is available from the homepage of Howard Bondell (see `http://www4.stat.ncsu.edu/~bondell/Software/NoCross/NoCrossQuant.R`, last date of access January 10, 2014). The programme for fitting noncrossing expectiles is part of the R package `expectreg` by Sobotka, Schnabel, and Schulze Waltrup (2013).

Conclusion: 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 direction of α.

3.4 Extensions and Examples

3.4.1 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 3.3.2, 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, Wand, and Carroll (2003, 2009). In other words we supplement (3.20) by the quadratic penalty

$$\lambda_u \boldsymbol{u}' \boldsymbol{D}' \boldsymbol{D} \boldsymbol{u}$$

where $\boldsymbol{D}'\boldsymbol{D}$ is an appropriately chosen penalty matrix and λ_u is the smoothing parameter chosen data driven. We give an example in the next subsection. For a specific value of α this has been proposed in Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013) for expectile smoothing or in Bollaerts, Eilers, and Aerts (2006) for quantile smoothing where the latter use a different penalization. The smoothing parameter λ_u can be chosen by asymmetric cross-validation or the Schall algorithm for mixed models as described in Schnabel and Eilers (2009b).

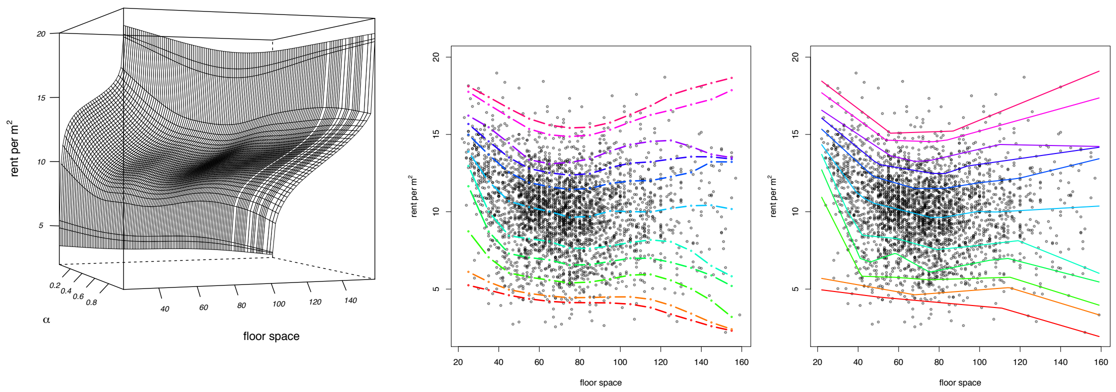


Figure 3.6: Expectile sheet for the Munich rent data (left), quantiles based on expectiles, calculated for certain values of x, which are visualized by points (middle) and non-crossing quantile smoothing splines (right). Quantiles are calculated for $\alpha = 0.01, 0.02, 0.05, 0.10, 0.20, 0.50, 0.80, 0.90, 0.95, 0.98, 0.99$.

3.4.2 Rent Index of Munich

To see how the method performs in practice, we again take a look at the Munich rental data from Section 3.2.4. As 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 nonparametric expectile regression as described in Sections 3.3.2 and 3.4.1. Our underlying model is given with the expectile sheet $m(\alpha, \text{living space})$. The smoothing parameter λ_u was chosen automatically by using the Schall algorithm (see Schnabel and Eilers, 2009b). The sheet resulting from the estimation procedure is shown in Figure 3.6. As one can see there, 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 3.2 and obtain the mid-panel of Figure 3.6. 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 meter as the apartment size grows. This continues till 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 our expectile based analysis we apply smooth spline based quantile fitting as described in Koenker, Ng, and Portnoy (1994) and implemented in the R package `quantreg` by Koenker (2013b). Here, α is kept fixed and smoothing is carried out separately for each value of α over the covariate only. Here non-crossing of quantile curves is not guaranteed. The resulting fit is shown in the right part of Figure 3.6. 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 quantile smoothing we decided to pick a smoothing parameter which results in a smoothness comparable to the amount of smoothing mirrored in the second panel of Figure 3.6.

3.4.3 Expected Shortfall

Investment risks are frequently measured using the expected shortfall (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 therefore be rather inefficient. Taylor (2008) presents a possibility to estimate the expected shortfall using expectiles and their connection to quantiles.

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

$$\frac{1-2\tilde{\alpha}}{\tilde{\alpha}} E[(Y - m_{\tilde{\alpha}}) I(Y < m_{\tilde{\alpha}})] = m_{\tilde{\alpha}} - E(Y) \tag{3.21}$$

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 equation (3.21) and using the fact $F(m_{\tilde{\alpha}}) = \alpha$ leads to

$$ES_{\text{low}}(\alpha) := E(Y|Y < q_\alpha) = \Big(1 + \frac{\tilde{\alpha}}{(1-2\tilde{\alpha})\alpha}\Big) m_{\tilde{\alpha}} - \frac{\tilde{\alpha}}{(1-2\tilde{\alpha})\alpha} m_{0.5} \tag{3.22}$$

for the lower tail of F. Depending on whether the random variable describes a win or a loss, we define the expected shortfall for the upper tail as $ES_{\text{up}}(\alpha) = E(Y|Y > q_{1-\alpha})$. 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 3.2.1 and estimate the distribution function $\hat{F}_m(.)$ from expectiles. As introduced in Section 3.2.1, we estimate a dense set of expectiles (i.e. we set $\alpha_l = 0.0005, 0.001, 0.005, 0.01, 0.02, \ldots, 0.98, 0.99, 0.995, 0.999, 0.9995$)

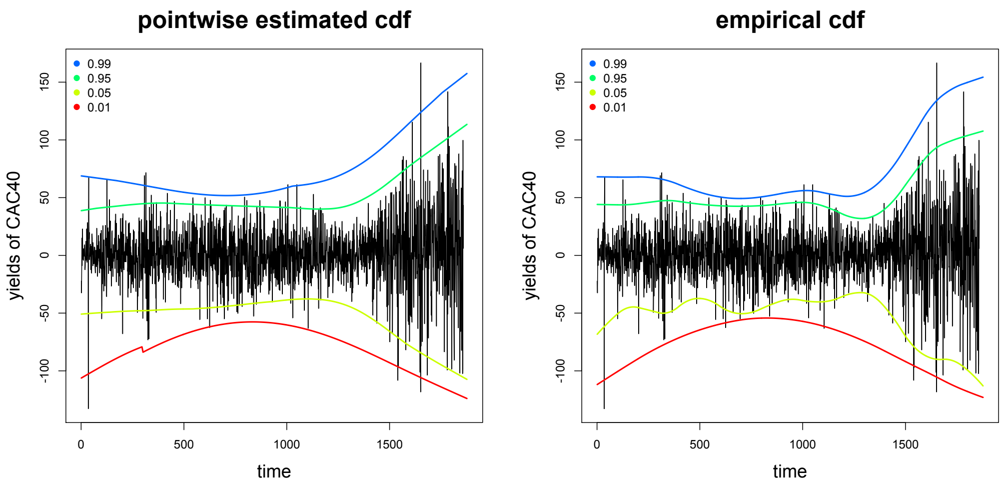


Figure 3.7: Estimated expected shortfall of the CAC40 yields for $\alpha = 0.01, 0.05, 0.95, 0.99$. Daily data from 1991 to 1998. Results with pointwise estimated distribution on the left, with empirical distribution function on the right.

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 expected shortfall explicitly for certain values of α, e.g. $\alpha = 0.01, 0.05, 0.95, 0.99$.

We apply the idea and estimate the expected shortfall 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 covariate influencing the expected shortfall. 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 $\lambda_u \boldsymbol{u}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{u}$ to (3.20), where $\boldsymbol{D}\boldsymbol{u}$ has rows $u_{lk} - u_{lk-1}$ for $l = 1, \ldots, L$ and $k = 1, \ldots, K$. The optimal smoothing parameter λ_u is chosen via asymmetric cross-validation, see Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013) for a more extensive description. Next, we apply the algorithm presented in Section 3.2 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 (3.22) for each point in time, that is sequentially observation by observation which gathers information about the changes in the distribution.

The result of the estimation is presented in Figure 3.7 (left part). As comparison we also fit the expected shortfall based on the empirical distribution as suggested by Taylor (2008) (right part of Figure 3.7). As can be seen, the volatility of the data is captured by the curves of the expected shortfall, 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, Yeh, and Hsu, 2009) that are able to account for the autocorrelation in the data with the methods introduced in this paper. Still, the example shows that an improvement in expected shortfall estimation is possible when using the efficient distribution estimation introduced in Section 3.2.1.

3.5 Discussion

In this paper 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 neighboring curves is an issue, we proposed a method to circumvent this problem. All methods regarding expectiles which were described in detail in this paper 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 expected shortfall. 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. Overall, the L_2 foundation of expectiles is helpful, as it allows to borrow penalty ideas from spline estimation and, of course, other extensions from the regression framework are possible as well. 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 expectiles provide an interesting and worthwhile alternative to the well established quantile regression.

Chapter 4

Non-crossing Expectile Smoothing for Panel Data using Penalized Splines

This chapter is developed in joint work with my supervisor Göran Kauermann. Göran Kauermann and Linda Schulze Waltrup developed the concept of non-crossing additive expectile regression which allows for the incorporation of a random intercept. Linda Schulze Waltrup implemented the method and explored the numerical performance. The analysis of the GSOEP data was carried out by Linda Schulze Waltrup. Both authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript.

Abstract

Expectile regression is a topic which became popular in the last years. It includes ordinary mean regression as special case but is more general as it offers the possibility to also model non-central parts of a distribution. Additive semiparametric expectile models recently have been developed and it is easy to perform flexible expectile estimation with modern software like R. Usually one estimates a whole set of expectile curves which, in theory, should not cross each other. In practice, however, crossing of neighboring expectile functions may occur. We discuss non-crossing expectile curves and propose a method to circumvent crossings of neighboring functions. We also allow for clustered observations, i.e. repeated measurements taken at the same individual. To accommodate the resulting dependence structure among the observations we extend semiparametric expectile estimation to semiparametric random effect expectile estimation. We apply our methods to panel data from the German Socio-Economic Panel.

4.1 Introduction

In the regression framework, one aims to estimate the expected value of a response in dependency on a set of covariates. Besides the dominant field of mean regression there are other, more general regression approaches. Expectile regression for example contains mean regression as special case but also offers the possibility to estimate tail expectations. Expectiles were introduced by Aigner, Amemiya, and Poirier (1976) and Newey and Powell (1987). They appear as alternative to the dominating quantiles (see Koenker and Bassett, 1978 or Koenker, 2005) and have seen some increasing interest in the last years. The two methods, i.e. quantiles and expectiles, both uniquely define a distribution function and hence allow for regression modeling beyond the mean. Jones (1994) and Yao and Tong (1996) derive the mathematical relation between quantiles and expectiles, see also De Rossi and Harvey, 2009. A general comparison of the two approaches is provided in Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014).

Expectiles can be fitted with asymmetric, iterated weighted least squares (Newey and Powell, 1987) which mirrors a L_2 distance measure like in the normal distribution. This resemblance has the great advantage that it allows to extend expectile regression in the same way as normal regression models have been generalized in the last decades. Schnabel and Eilers (2009b) propose expectile smoothing with penalized splines, see also Pratesi, Ranalli, and Salvati (2009). Sobotka and Kneib (2012) use spatial smoothing in combination with expectiles. Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013) provide asymptotic properties while Guo and Härdle (2013) propose simultaneous confidence bands for expectile curves. Taylor (2008) uses expectiles to calculate the expected shortfall. A general overview over expectile regression is found in Kneib (2013), see also De Rossi and Harvey (2009). In this paper we extend smooth expectile regression to cope with clustered observations. To be specific, we look at panel data giving the income of an individual over time and age. The data trace from the German Socio Economic panel (Wagner, Frick, and Schupp, 2007) and we have clustered panel based observations. For each individual in the study we include a random effect yielding a linear mixed model which is extended to expectiles.

Quantile regression for longitudinal, clustered data has been proposed in Tang and Leng (2011), see also Leng and Zhang, 2014 or Koenker, 2004. The latter uses random effects to estimate a single quantile. We extend this to expectiles, but estimate the expectiles for a whole range of asymmetry values $\alpha \in (0, 1)$, which will be labelled as an expectile sheet in the paper. The modeling of quantile and expectile sheets has been proposed in Schnabel and Eilers (2013b). The idea is, instead of estimating quantiles (or expectiles) for a specific asymmetry value α, one fits a complete surface, the so called quantile (or expectile) sheet giving the value of the α-th quantile conditional on the covariate value x, say. The sheet is a bivariate function which needs to be estimated from the data subject to some monotonicity condition to avoid crossing quantiles (see also Koenker and Ng, 2005). We employ the idea to estimate expectile sheets which depend additively on the covariates in the model.

The paper is organized as follows. In Section 4.2 we develop the theory of non-crossing random effect expectile regression. We start in Subsections 4.2.1 and 4.2.2 by describing spline based expectile estimation of expectile sheets. Here we concentrate on one covariate of influence. In Subsection 4.2.3 we extend the model to numerous covariates. As a last step in Subsections 4.2.4 and 4.2.5 a random intercept will be introduced. Then we take a turn to Section 4.3.1 where we apply our methods to the data from the German Socio-Economic Panel. A conclusion finalizes the paper.

4.2 Modeling Smooth Expectiles

4.2.1 Spline Expectile Estimation

Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$ denote a set of longitudinal or panel observations, respectively, taken on cluster or individual i. In our example y_{ij} gives the monthly income of an individual i in panel wave j. With $\boldsymbol{X}_i = (\boldsymbol{x}'_{i1}, \ldots, \boldsymbol{x}'_{in_i})'$ we denote the matching matrix of covariates, where $\boldsymbol{x}_{ij} = (x_{1ij}, \ldots, x_{dij})'$ is a d dimensional vector, $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. For ease of presentation, we first assume that $d = 1$ and let x_{1ij} be metrically scaled. Moreover, we will ignore for the moment that observations taken at the same individual or at the same cluster are dependent. This corresponds in principle to unclustered data, i.e. taking $n_j \equiv 1$, but it would be conceptually and notationally misleading to restrict the subsequent presentation to this special case. Accordingly for now we assume that $\boldsymbol{y}_i$ is a vector and $\boldsymbol{X}_i$ is a $n_i \times 1$ dimensional matrix. Our interest is in the estimation of the expectile sheet $m(\alpha, x_1)$ which is assumed to be a smooth function in both, covariate x_1 and asymmetry value $\alpha \in (0, 1)$. Defining the residual with $\epsilon = y - m(\alpha, x_1)$, the expectile sheet $m(\alpha, x_1)$ is implicitly defined through

$$m(\alpha, x_1) = \arg\min E\left(w_\alpha(\epsilon)\epsilon^2\right), \tag{4.1}$$

where

$$w_\alpha(\epsilon) = \begin{cases} \alpha & \text{for } \epsilon \geq 0 \\ 1 - \alpha & \text{for } \epsilon < 0. \end{cases} \tag{4.2}$$

The expectation in (4.1) is thereby taken with respect to the conditional distribution of y given x_1. Note that for $\alpha = 0.5$ the classical smoothing model results, i.e. $m(0.5, x_1) = E(y|x_1)$. Moreover, it is easy to see that $m(\alpha, x_1)$ is monotone in α, that is $m(\alpha_1, x_1) \leq m(\alpha_2, x_1)$ for $\alpha_1 \leq \alpha_2$.

Our presentation of modeling and fitting expectile sheet $m(\alpha, x_1)$ subsequently follows three steps. First, we need to guarantee that the fitted function is monotone in α, i.e. we need to fit non-crossing expectile functions. As a second step, we extend the model to allow for multiple covariates. Finally, as third step, the model needs to be generalized to allow for clustered observations.

We start by providing a numerical procedure for non-crossing expectile estimation, see also Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014). To do so, we replace function $m(\alpha, x_1)$ by a spline basis representation following the theory and practice of penalized spline smoothing, see Ruppert, Wand, and Carroll (2003, 2009). To be more specific, let $\boldsymbol{B}_{(1)}(x_1) = (B_1(x_1), \ldots, B_K(x_1))$ be a B-spline basis of polynomial order p built upon the knots $\kappa_{11}, \ldots, \kappa_{1(K-p+1)}$, which cover the (observed) range of covariates x_1. Moreover, let $\tilde{\boldsymbol{B}}_{(0)}(p) = (\tilde{B}_1(\alpha), \ldots, \tilde{B}_M(\alpha))$ be a linear B-spline basis built upon the knots $0 < \tilde{\alpha}_1 < \tilde{\alpha}_2 < \cdots < \tilde{\alpha}_M < 1$. We now approximate $m(\alpha, x_1)$ through the tensor product

$$m(\alpha, x_1) = \sum_{k=1}^{K} \sum_{m=1}^{M} \tilde{B}_m(\alpha) B_k(x_1) u_{km} \tag{4.3}$$

$$= (\tilde{\boldsymbol{B}}_{(0)}(\alpha) \otimes \boldsymbol{B}_{(1)}(x_1))\boldsymbol{u} =: \boldsymbol{B}(p, x_1)\boldsymbol{u}, \tag{4.4}$$

where $\boldsymbol{u} = (u_{11}, \ldots, u_{1K}, u_{21}, \ldots, u_{MK})'$. The use of a linear B-spline in direction of α corresponds to estimating a set of expectiles for $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_M$ and use a simple linear interpolation for the calculation of expectiles for any $\alpha \in [\tilde{\alpha}_m, \tilde{\alpha}_{m+1}]$. This justifies the use of a linear basis and mirrors the role of the number of knots M. The larger we choose M, the finer the grid for linear interpolation. Apparently, if we choose M too large we will face a numerical limit. Other than that the selection of M plays a minor role. Note that the selection of the number of knots in x direction, that is K, plays a different role and more discussion is required, which we postpone for the moment. Replacing now the expectile definition (4.1) by its empirical version provides the weighted least squares estimate for coefficient vector $\boldsymbol{u}$ through

$$\hat{\boldsymbol{u}} = \arg\min \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{m=1}^{M} w_{\alpha_m} \left(y_{ij} - \boldsymbol{B}(\alpha_m, x_{1ij})\boldsymbol{u}\right) \left\{y_{ij} - \boldsymbol{B}(\alpha_m, x_{1ij})\boldsymbol{u}\right\}^2 \tag{4.5}$$

with $w_{\alpha_m}(.)$ as defined subsequent to (4.2). To continue it is notationally advisable to rewrite (4.5) in a more comprehensive matrix form. Let therefore $\boldsymbol{y} = \boldsymbol{1}'_M \otimes (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n)$ and $\boldsymbol{B}$ denoting the matrix having rows $\boldsymbol{B}(\alpha_m, x_{1ij})$ for $m = 1, \ldots, M$; $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Finally, we denote with $\boldsymbol{W}$ the diagonal weight matrix having $w_{\alpha_m}(y_{ij} - m(\alpha_m, x_{1ij}))$ on its diagonal. Then (4.5) can be written as

$$\hat{\boldsymbol{u}} = \arg\min(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u}) =: \arg\min Q(\boldsymbol{u}) \tag{4.6}$$

with $Q(\boldsymbol{u})$ denoting the quadratic form in (4.6). Based on (4.6) we can easily derive an estimation routine using iterative weighted least squares. That is, we ignore the dependence of weight matrix $\boldsymbol{W}$ on $\boldsymbol{u}$ and solve (4.6). Fixing the weights, i.e. considering weight matrix $\boldsymbol{W}$ as given, provides the minimum of (4.6) as

$$\hat{\boldsymbol{u}} = (\boldsymbol{B}'\boldsymbol{W}\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{W}\boldsymbol{y}\ . \tag{4.7}$$

With this tentative estimate $\hat{\boldsymbol{u}}$ we now recalculate the weight matrix $\boldsymbol{W}$ and recalculate (4.7). These two steps are repeated until convergence.

The resulting fit $\boldsymbol{B}\hat{\boldsymbol{u}}$ may not necessarily be a valid expectile sheet since monotonicity in α is not guaranteed. We therefore need to impose constraints on vector $\boldsymbol{u}$ to obtain non-crossing expectiles. Looking at (4.3) and taking advantage of the linear B-spline structure for $\tilde{B}_m(\alpha)$, we obtain non-crossing expectile functions as long as

$$\boldsymbol{B}_{(1)}(x_1)(\boldsymbol{u}_{m\cdot} - \boldsymbol{u}_{m+1\cdot}) \leq 0 \quad \text{for } m = 1, \ldots, M-1 \tag{4.8}$$

is guaranteed, where $\boldsymbol{u}_{m\cdot} = (u_{m1}, \ldots, u_{mk})$ is the corresponding subvector of $\boldsymbol{u}$. Apparently, (4.8) gives a set of simple linear constraints which are easily accommodated in the estimation routine using e.g. quadratic programming for minimizing (4.6) subject to (4.8).

4.2.2 Penalization

We return now to the role of spline dimension K, i.e. the spline basis in direction of covariate x_1. We observe that the fitted curve becomes wiggled if K is too large, i.e. the model is overparameterized and parameter estimates become unstable. We will therefore introduce a penalty term which compromizes numerical stability with modeling flexibility. Following the idea of penalized spline fitting in the line of Eilers and Marx (1996) we penalize r-th order differences of the neighboring elements of subvector $\boldsymbol{u}_{m\cdot} = (u_{1m}, \ldots, u_{Km})^T$ with $m = 1, \ldots, M$. Let therefore $\boldsymbol{D}_K$ be the $(K-r) \times K$ dimensional difference matrix, then the penalty on $\boldsymbol{u}_{m\cdot}$ is defined through $\lambda \boldsymbol{u}'_{m\cdot}\boldsymbol{K}_K\boldsymbol{u}_{m\cdot}$ where $\boldsymbol{K}_K = \boldsymbol{D}'_K\boldsymbol{D}_K$ and λ is the penalty parameter. This can be written for the entire vector $\boldsymbol{u}$ through

$$\lambda \boldsymbol{u}'\boldsymbol{K}\boldsymbol{u},$$

where $\boldsymbol{K} = \boldsymbol{I}_M \otimes \boldsymbol{K}_K$. The penalty parameter λ steers the smoothness of the fitted expectile functions $\hat{m}(\alpha, x)$ where $\lambda \to \infty$ leads to a smooth polynomial fit. Our intention is now to minimize the constrained penalized weighted least squares

$$Q(\boldsymbol{u}) + \frac{1}{2}\lambda \boldsymbol{u}'\boldsymbol{K}\boldsymbol{u} \text{ subject to } \boldsymbol{A}\boldsymbol{u} \geq \boldsymbol{0}, \tag{4.9}$$

where the linear constraints matrix $\boldsymbol{A}$ is easily constructed from (4.8).

4.2.3 Additive Expectile Estimation

We now extend the above estimation routine to incorporate multiple covariates in a structured additive framework. Let therefore $\boldsymbol{x}_{ij}$ be the d dimensional covariate vector ordered such that $x_{1ij}, \ldots, x_{\tilde{d}ij}$ are (binary) factorial covariates and $x_{\tilde{d}+1ij}, \ldots, x_{dij}$ are metrically scaled, respectively with $1 \leq \tilde{d} \leq d$. We model the expectile function additively and write the α-th expectile $m(\alpha, \boldsymbol{x}_{ij})$ as

$$m(\alpha, \boldsymbol{x}_{ij}) = \sum_{h=0}^{d} m_h(\alpha, x_{hij}) = \sum_{h=0}^{\tilde{d}} x_{hij}\beta_h(\alpha) + \sum_{h=\tilde{d}+1}^{d} m_h(\alpha, x_{hij}), \tag{4.10}$$

where $\beta_h(\alpha)$ is a smooth function in α and $x_{0ij} \equiv 1$ is the intercept. Apparently, model (4.10) needs identifiability constraints, a problem which will be solved automatically within the estimation procedure as explained below. Note that the expectile approach extends traditional additive models (for which $\alpha = 0.5$) as proposed for instance in Hastie and Tibshirani (1990) (see also Wood, 2006). Like before, we replace the smooth components by B-splines and we define $\boldsymbol{B}_{(h)}(\alpha, x_{hij}) = x_{hij}\tilde{\boldsymbol{B}}_{(0)}(\alpha)$ for $h = 0, \ldots, \tilde{d}$ and set

$$\boldsymbol{B}_{(h)}(\alpha, x_{hij}) = \tilde{\boldsymbol{B}}_{(0)}(\alpha) \otimes \boldsymbol{B}_{(h)}(x_{hij})$$

for $h = \tilde{d}+1, \ldots, d$, where $\boldsymbol{B}_{(h)}(.)$ is a K_h dimensional B-spline basis of order p built upon the knots $\kappa_{h1}, \ldots, \kappa_{h(K_h-p+1)}$ covering the (observed) values of x_h. As before $\tilde{\boldsymbol{B}}_{(0)}(\alpha)$ is the linear B-spline basis. This allows to rewrite (4.10) to

$$m(\alpha, \boldsymbol{x}_{ij}) = \sum_{h=0}^{d} \boldsymbol{B}_{(h)}(\alpha, x_{hij})\boldsymbol{u}_h \ .$$

For $\tilde{d}+1 \leq h \leq d$ subvector $\boldsymbol{u}_h$ decomposes to components u_{hkm} with $k = 1, \ldots, K_h$ and $m = 1, \ldots, M$. This is completely analogous to the previous section. For $h = 0, \ldots, \tilde{d}$ we model the expectile dependent slope parameter $\beta_h(\alpha)$ in (4.10) as linear interpolation of the expectiles for the knot locations $\alpha_1, \ldots, \alpha_m$. Therefore for $0 \leq h \leq \tilde{d}$ subvector $\boldsymbol{u}_h$ decomposes to components u_{h1m} with $m = 1, \ldots, M$. The coherent setting (4.10) allows now to directly extend the estimation routine from the previous section. Let therefore $\boldsymbol{u} = (\boldsymbol{u}_0', \ldots, \boldsymbol{u}_d')'$ and $\boldsymbol{B} = (\boldsymbol{B}_{(0)}, \ldots, \boldsymbol{B}_{(d)})$ where $\boldsymbol{B}_{(h)}$ is the $(M \cdot \sum_{i=1}^{n} n_i) \times (K_h \cdot M)$ dimensional basis with rows $\boldsymbol{B}_{(h)}(\alpha_m, x_{hij}), m = 1, \ldots, M,\ i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. With $Q(\boldsymbol{u})$ we define

$$Q(\boldsymbol{u}) = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u}),$$

where $\boldsymbol{W}$ is a diagonal weight matrix as defined in the previous section. Then we intend to minimize the constrained penalized least squares

$$\arg\min \{Q(\boldsymbol{u}) + \sum_{h=\tilde{d}+1}^{d} \lambda_h \boldsymbol{u}_h' \boldsymbol{K}_h \boldsymbol{u}_h\} \text{ subject to } \boldsymbol{A}\boldsymbol{u} \geq \boldsymbol{0}, \tag{4.11}$$

where $\boldsymbol{K}_h$ are penalty matrices constructed by r-th order differences and λ_h as penalty parameter steering the smoothness of component $m_h(\alpha, x_h)$ with respect to x_h. The constraints matrix $\boldsymbol{A}$ can be defined as follows. As in the one-dimensional case, we want to estimate monotone (non-crossing) expectiles. That is, we need

$$m(\alpha_1, \boldsymbol{x}_{ij}) \geq m(\alpha_2, \boldsymbol{x}_{ij}) \tag{4.12}$$

for $\alpha_1 \geq \alpha_2$. Note that (4.12) means, we have to ensure that

$$\sum_{h=0}^{d} m_h(\alpha, x_{hij}) - m_h(\tilde{\alpha}, x_{hij}) \geq 0 \tag{4.13}$$

holds for all $i = 1, \ldots, n$. Remembering that $m_h(\alpha, x_{hij}) = \boldsymbol{B}_h(\alpha_m, x_{hij})\hat{\boldsymbol{u}}_h$, we again can rewrite (4.13) as linear constraints leading to

$$\boldsymbol{A}\boldsymbol{u} \geq \boldsymbol{0}. \tag{4.14}$$

In principle, one can now fit an additive non-crossing semiparametric expectile model. But, as mentioned above, we have to ensure identifiability of the model and determine the smoothing parameters appropriately. To solve both problems, we make use of a mixed model representation of penalized splines as described for regular smoothing, i.e. for $\alpha = 0.5$ in Ruppert, Wand, and Carroll (2003) and Fahrmeir, Kneib, Lang, and Marx (2013). Before going into details in this respect we incorporate random effects to accommodate the clustered structure of the observations.

4.2.4 Random Effect Expectile Estimation

We extend the above model by taking into account that observations $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})'$ trace from the same individual or cluster, respectively. Following the idea of linear mixed models we include an individual random effect γ_i in the model which provides a linear shift for all observations from the same individual, $i = 1, \ldots, n$. Denoting with $\epsilon_{ij} = y_{ij} - \gamma_i - m(\alpha, \boldsymbol{x}_{ij})$ the residual belonging to observation j of individual i we define the α-th expectile function through

$$m(\alpha, \boldsymbol{x}_{ij}) = \arg\min E\big(w_\alpha(\epsilon_{ij})\epsilon_{ij}^2\big),$$

where the expectation is taken with respect to y_{ij} but conditional on γ_i and covariates $\boldsymbol{x}_{ij}$. We assume that γ_i is random and make use of the prior model

$$\gamma_i \sim N(0, \sigma_\gamma^2) \text{ i.i.d. for } i = 1, \ldots, n. \tag{4.15}$$

To incorporate prior distribution (4.15) in the estimation routine, we reformulate (4.15) as penalty term in the estimating steps below, comparable to the penalization of the spline coefficients. Let therefore $\boldsymbol{U}$ be the indicator matrix relating the observations from the i-th individual in vector $\boldsymbol{y} = \boldsymbol{1}'_M \otimes (\boldsymbol{y}'_1, \ldots, \boldsymbol{y}'_n)$ to the i-th individual effect γ_i. We extend the quadratic form $Q(\boldsymbol{u})$ to

$$Q(\boldsymbol{u}) = (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u} - \boldsymbol{U}\boldsymbol{\gamma})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{u} - \boldsymbol{U}\boldsymbol{\gamma}) \tag{4.16}$$

with obvious definition for weight matrix $\boldsymbol{W}$. The task is now to minimize the constrained least squares criterion

$$\arg\min \{Q(\boldsymbol{u}) + \sum_{h=\tilde{d}+1}^{d} \lambda_h \boldsymbol{u}_h' \boldsymbol{K}_h \boldsymbol{u}_h + \lambda_\gamma \boldsymbol{\gamma}'\boldsymbol{\gamma}\} \text{ subject to } \boldsymbol{A}\boldsymbol{u} \geq \boldsymbol{0},$$

where $\lambda_\gamma = \frac{\sigma_\epsilon^2}{\sigma_\gamma^2}$.

4.2.5 Mixed Model Representation

Looking at the quadratic form $Q(\boldsymbol{u})$ in (4.16) and ignoring for the moment that $\boldsymbol{W}$ depends on the unknown parameter, we may understand $Q(\boldsymbol{u})$ to result from the model

$$\begin{aligned} \boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\gamma} &\sim N(\boldsymbol{B}\boldsymbol{u} + \boldsymbol{U}\boldsymbol{\gamma}, \sigma_\epsilon^2 \boldsymbol{W}^{-1}) \\ \boldsymbol{u}_h &\sim N(0, \sigma_h^2 \boldsymbol{K}_h^-) \text{ i.i.d. } h = \tilde{d}+1, \ldots, d \\ \boldsymbol{\gamma} &\sim N(0, \sigma_\gamma^2 \boldsymbol{I}_n) \end{aligned} \tag{4.17}$$

with $\sigma_h^2 = \frac{\sigma_\epsilon^2}{\lambda_h}$. This is in direct analogy to the link between spline fitting and linear mixed models as extensively discussed in Ruppert, Wand, and Carroll (2003) and Kauermann, Krivobokova, and Fahrmeir (2009). Formulation (4.17) allows to derive Maximum Likelihood estimates based on $\boldsymbol{y}$ for parameters $\sigma_\epsilon^2, \lambda_h$ and λ_γ. The numerical calculation has been proposed by Schall (1991), but we also refer to Schnabel and Eilers (2009b) who apply the idea to expectile smoothing. Assuming model (4.17), we have an (unconstrained) estimate of $\boldsymbol{v} = (\boldsymbol{u}', \boldsymbol{\gamma}')'$ through

$$\hat{\boldsymbol{v}} = \left(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{K}\right)^{-1} \boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y},$$

where $\boldsymbol{Z} = (\boldsymbol{B}, \boldsymbol{U})$ and $\boldsymbol{K} = \boldsymbol{K}(\lambda_h, h = \tilde{d}+1, \ldots, d, \lambda_\gamma)$ is the block diagonal matrix build from 0 matrices for entries corresponding to $\boldsymbol{u}_h, h = 0, \ldots, \tilde{d}$ and $\lambda_h \boldsymbol{K}_h$ for entries matching to $h = \tilde{d}+1, \ldots, d$ and finally $\lambda_\gamma \boldsymbol{I}_n$ on its diagonal. It is now not difficult to show that, assuming orthogonality of the basis functions, the Maximum Likelihood estimates for λ_h and λ_γ can be approximated through

$$\hat{\sigma}_h^2 = \frac{\hat{\boldsymbol{u}}_h' \boldsymbol{K}_h \hat{\boldsymbol{u}}_h}{\mathrm{df}_h(\lambda_h)}, \quad \hat{\sigma}_\gamma^2 = \frac{\hat{\boldsymbol{\gamma}}'\hat{\boldsymbol{\gamma}}}{\mathrm{df}_\gamma(\lambda_\gamma)} \quad \text{and} \quad \hat{\sigma}_\epsilon^2 = \frac{(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{v}})'\boldsymbol{W}(\boldsymbol{y} - \boldsymbol{Z}\hat{\boldsymbol{v}})}{M\sum_{i=1}^n n_i - \mathrm{df}}$$

where

$$\mathrm{df} = \mathrm{tr}\Big(\big(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{K}\big)^{-1}\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{Z}\Big) \tag{4.18}$$

and

$$\mathrm{df}_h = \mathrm{tr}\Big(\big(\tilde{\boldsymbol{Z}}'\boldsymbol{W}\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{K}}\big)^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{W}\tilde{\boldsymbol{Z}}\Big)_h \tag{4.19}$$

with $\tilde{\boldsymbol{Z}}$ and $\tilde{\boldsymbol{K}}$ consisting only of the random part of the corresponding matrices $\boldsymbol{Z}$ and $\boldsymbol{K}$, i.e. corresponding to the non-zero diagonal of $\boldsymbol{K}$ (and $\tilde{\boldsymbol{Z}}$ resulting from the mixed model representation). As before, subindex h in (4.19) means that we solely extract the columns belonging to the h-th covariate. Analogously $\mathrm{df}_\gamma(\lambda_\gamma)$ can be defined by extracting the columns of $\Big(\big(\tilde{\boldsymbol{Z}}'\boldsymbol{W}\tilde{\boldsymbol{Z}} + \tilde{\boldsymbol{K}}\big)^{-1}\tilde{\boldsymbol{Z}}'\boldsymbol{W}\tilde{\boldsymbol{Z}}\Big)$ belonging to the random intercept. The mixed model representation also implies that identifiability issues resulting from model (4.10) disappear since we may comprehend the priors on the parameters as Bayesian priors which regularize the estimation. In fact if $\sigma_h^2 \to 0$ the effect of a covariate is set to zero.

4.3 Example

4.3.1 Application to German Socio-Economic Panel Data

To show the applicability of our approach we consider data from the German Socio-Economic Panel (GSOEP, see www.diw.de/soep). The GSOEP is a detailed questionnaire starting in 1984 and still ongoing. We focus our attention on data from 2003 until 2012 and look at the monthly gross income of women living in the western part of Germany. We hope to explore more interesting patterns for women than for men and therefore decided to concentrate on women. Due to historical reasons there is still a gap between the amount of income in the western part of Germany and the amount of income in the eastern part of Germany (former GDR). Monthly gross income is reported on a yearly basis and inflation-adjusted with 2003 as reference year. We distinguish between three different levels of education measured with the “International Standard Classification of Education” (ISCED) which originally consists of seven levels. We use levels 0 to 2 as category “lower education”, levels 3 and 4 as “middle education” and 5 and 6 as “higher education”. Additionally we have age as covariate where we restrict our attention to women between 25 an 60.

As we are interested not only in mean income but also in higher and lower income we

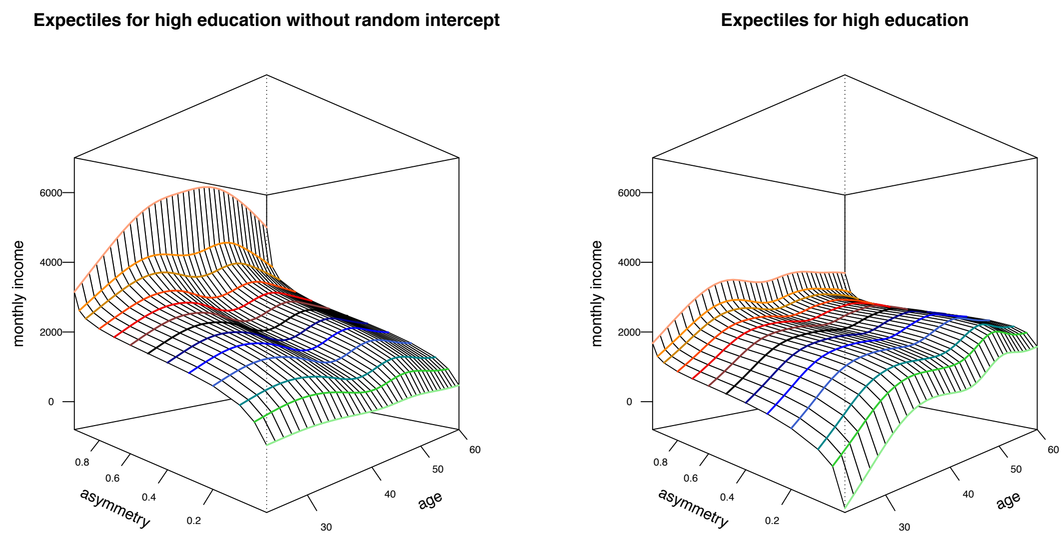


Figure 4.1: The relationship between adjusted income and age of highly educated women in western Germany during 2003 - 2012 using model (4.20) (left panel) and model (4.21) (right panel). Highlighted expectiles correspond to $\alpha = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.99$.

use the techniques described in the previous subsections to fit the expectile model

$$y_{ij} = m(\alpha, x_{1ij}) + \epsilon_{ij} \tag{4.20}$$

with adjusted gross income as variable of interest and age as explanatory variable. As described before, model (4.20) ignores the special dependence structure of the data. To account for repeated measurements we add a random shift to each women which leads to model

$$y_{ij} = m(\alpha, x_{1ij}) + \gamma_i + \epsilon_{ij}. \tag{4.21}$$

The models were fitted applying the methods described in the previous subsections. An implementation is available within the open source software R (see R package `expectreg`). In Figure 4.1 we see the results for highly educated women based on model (4.20) (left plot) and model (4.21) (right plot). A first conclusion from the two plots is that the random intercept captures a large amount of the variance, as the curves in the right panel lie more

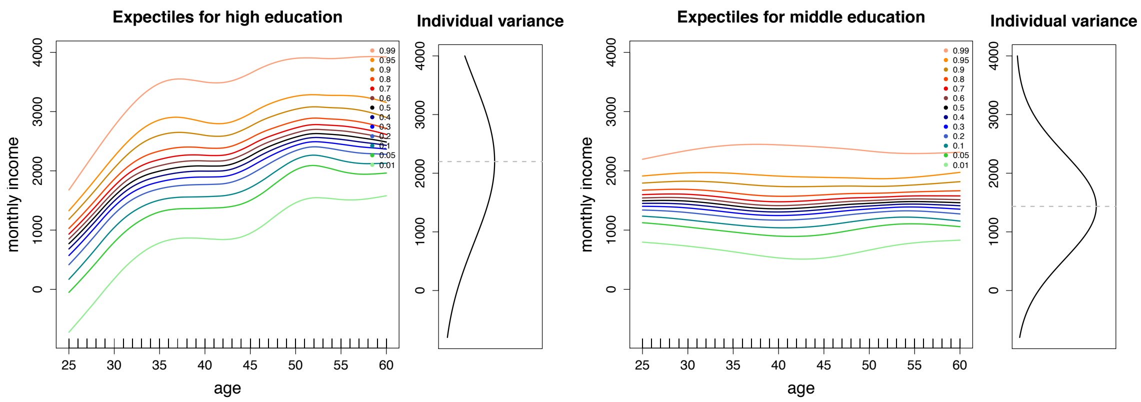


Figure 4.2: The relationship between adjusted income and age of highly educated women (left) and women of middle education (right) in western Germany during 2003 - 2012 using model (4.21). Expectiles correspond to $\alpha = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.99$. Additionally the individual variance in income is visualized.

close to each other than the curves in the left panel. The right panel shows that with increasing age also gross income increases. There seems to be phases in women's lives where income stagnates: approximately between 35 and 45 and for women older than 50. After taking the panel structure into account the behaviour of all expectile curves is similar which means that there appears only moderate difference between the relationship of income and age for women with high income (corresponding to higher expectile curves) and women with low income (represented by lower expectile curves). The standard deviations σ_ϵ and σ_γ are estimated automatically using the Schall algorithm (see Section 4.2.5). For the women of high income, σ_ϵ and $\sigma_{\gamma,\text{high}}$ are estimated as $\hat{\sigma}_\epsilon = 649.99$ and $\hat{\sigma}_{\gamma,\text{high}} = 1384.42$ which implies that there is a strong deviation of income between different women.

To get a better impression of the relationship between income and age we extract the expectile curves corresponding to $\alpha = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80,$ $0.90, 0.95, 0.99$ from the expectile sheet in Figure 4.1. The result can be seen in the left plot in Figure 4.2. We also visualize the fitted individual variation based on the normal prior of model (4.17). The right part of Figure 4.2 shows the relationship between income and age for women of middle education. We see that for women of middle education the relationship between age and income seems to be nearly linear or, more precisely about constant, with a slight cutback during the years from age 35 to 45. During that specific range of age we also saw the stagnation of income for women of high education. The variance in monthly income between women of middle education is estimated as $\hat{\sigma}_{\gamma,\text{middle}} = 900.71$ and therewith clearly smaller than the individual variation for high-educated women. Turning the attention to

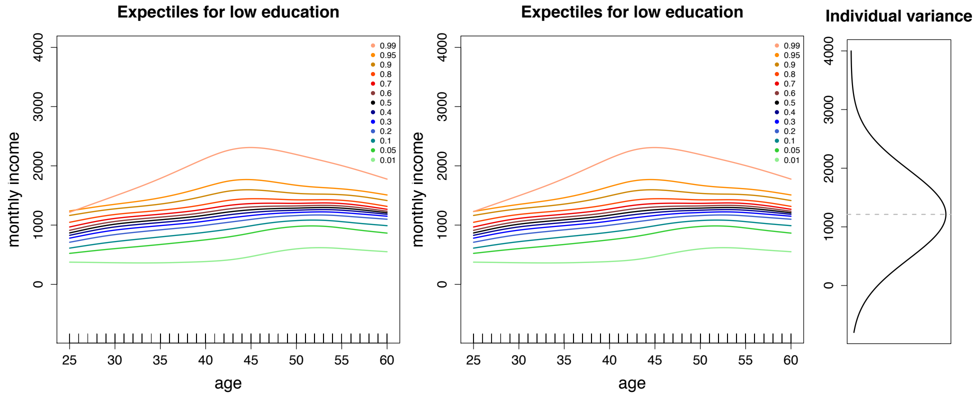


Figure 4.3: The relationship between adjusted income and age for women of low education in western Germany during 2003 - 2012 using model (4.21). Expectiles correspond to $\alpha = 0.01, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.99$. Additionally the individual variance in income is visualized. The first plot shows the unconstrained fit of model (4.21) where we observe a crossing between the 0.99 and 0.95 expectile. In the second plot the constrained, non-crossing fit of model (4.21) can be seen.

women of lower education we see from Figure 4.3 that there seems to be an increase in monthly income during the ages of 25 to 50. Afterwards the income is stagnating or slightly falling, a feature which was also present for highly educated women. The variance in monthly income between women of low education is estimated as before automatically and constitutes $\hat{\sigma}_{\gamma,\mathrm{low}} = 762.15$. For both women of high and middle education we see a stagnation or decline in monthly income for the ages from 35 to 45 which may be explained by a phase in women's live where raising of children comes to the fore. The variation in income between women is the highest for highly educated women and declines with decreasing education. We also looked at log(monthly income) as response which led to similar conclusions.

4.3.2 Bootstrap Confidence Bands

In a next step we will examine the certainty of the expectile curves shown in the previous figures. Therefore we run a nonparametric bootstrap. A detailed introduction into the bootstrap can be found in Efron and Tibshirani (1993). To do so, we draw with replacement n women from the set of n women and estimate model (4.21) for the corresponding new data set. To simplify numerics we regard λ as fixed, that is, we ignore the variability induced through the estimation of λ. In principle, we could assess the uncertainty via asymptotic confidence intervals as described in Sobotka, Kauermann, Schulze Waltrup, and

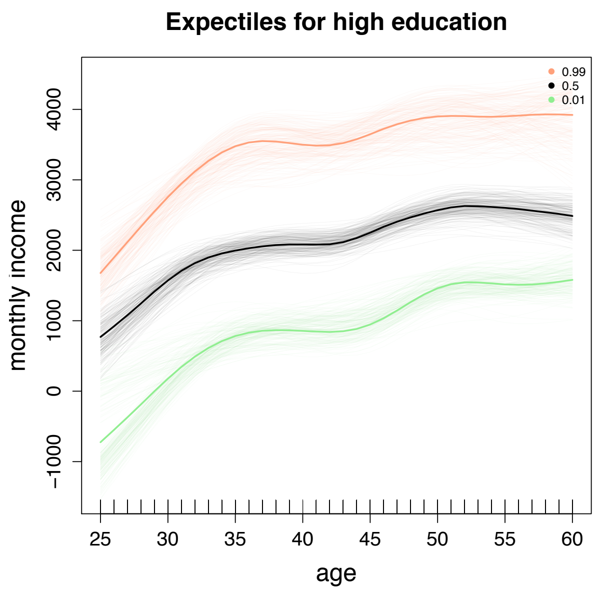


Figure 4.4: The relationship between adjusted income and age for women of high education using model (4.21). The shaded curves result from a nonparametric bootstrap with 200 bootstrap samples generated.

Kneib (2013). Here, however, we have $n_{\text{high}} = 293, n_{\text{middle}} = 636$ and $n_{\text{low}} = 192$ which is slightly too low to recommend the use of the asymptotic confidence intervals. Figure 4.4 shows the expectile curves for the highly educated women (see Figure 4.2) for $\alpha = 0.01, 0.5, 0.99$. The variability of the fit can be evaluated by looking at the shaded lines around the 0.01, 0.5 and 0.99 expectile, respectively. Each shaded line is resulting from one of the 200 bootstrap samples. We see that even after taking the variation within the sampling process into account there is a strong non linear behaviour in the relationship between monthly income and age for highly educated women. Also the change between periods of rise in income and stagnation in income remains present.

4.3.3 Additive Expectile Estimation

To further illustrate the applicability of our method we extend model (4.21) and allow for additional covariates. Apparently numerous variables may potentially influence the monthly gross income of women. But for illustrative purposes we restrict our attention here to the employment time that the women spent with the company she is working at. This leads to

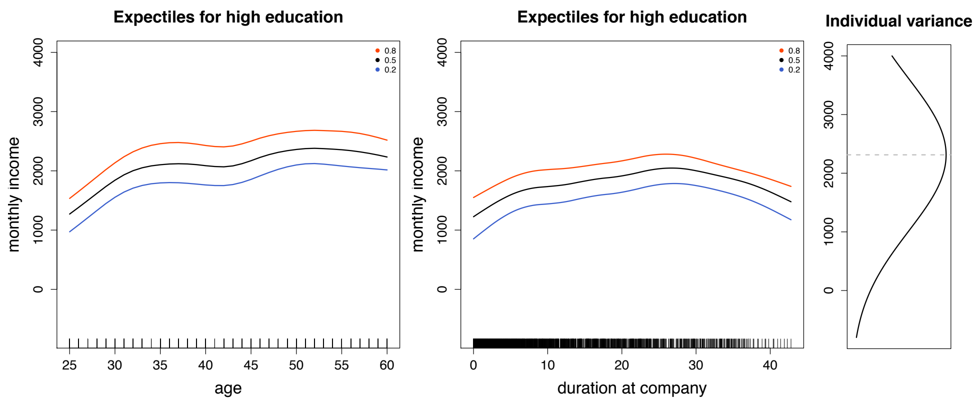


Figure 4.5: The relationship between adjusted income, age (left plot) and duration time at company (middle plot) for women of high education in western Germany during 2003 - 2012 for model (4.22). Expectiles correspond to $\alpha = 0.20, 0.50, 0.80$. Additionally the individual variance in income is visualized (right plot).

the additive expectile model

$$y_{ij} = m(\alpha, \boldsymbol{x}_{ij}) + \gamma_i + \epsilon_{ij} = m(\alpha, x_{1ij}) + m(p, x_{2ij}) + \gamma_i + \epsilon_{ij}, \tag{4.22}$$

where y denotes adjusted gross income and x_1 is the age of the female, as before, and x_2 is the employment time at the current employer. As before γ_i allows for an individual shift. Covariates age and duration are modeled as smooth functions and the smoothing parameter is selected as described in Subsection 4.2.5. In Figure 4.5 we see the result for high-educated women. The variances σ^2_ϵ and σ^2_γ are estimated as $\hat{\sigma}^2_\epsilon = 624.07$ and $\hat{\sigma}^2_{\gamma,\text{high}} = 1299.96$ which is slightly lower when only regarding age as covariate. We see that the relationship between monthly income and duration time is positive for a duration up to approximately 25 years but is decreasing afterwards. The relationship between income and age remains nearly the same as induced by model (4.21).

4.4 Conclusion

In this paper we extended expectile sheet estimation towards clustered observations. We included random intercepts which in combination with spline based fitting allowed for coherent estimation. We demonstrated the applicability with a typical data example. All in all the paper demonstrates that the use of expectiles allows to extend the model class in the same way as mean regression models have been extended in the last decades. We also addressed the problem of crossing expectile curves and made a suggestion how to circumvent crossings of neighboring expectile functions. The methods described in this paper are implemented in the open source software R (see R Core Team, 2014) and can be found in the R package `expectreg` by Sobotka, Schnabel, and Schulze Waltrup (2013).

Chapter 5

A Short Note on Quantile and Expectile Estimation in Unequal Probability Samples

This chapter is developed in joint work with my supervisor Göran Kauermann. Linda Schulze Waltrup and Göran Kauermann developed in cooperation the connection between expectiles and quantiles. Göran Kauermann proposed the estimation of weighted expectiles within the frameweork of unequal probability sampling. Linda Schulze Waltrup proposed and explored the use of the plug-in distribution estimator. The simulation study was run by Linda Schulze Waltrup. Both authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript.

Abstract

The estimation of quantiles is an important topic not only in the regression framework, but also in sampling theory. A natural alternative or addition to quantiles are expectiles. Expectiles as a generalization of the mean have become popular during the last years as they not only give a more detailed picture of the data than the ordinary mean, but also can serve as a basis to calculate quantiles by using their close relationship. We show, how to estimate expectiles under sampling with unequal probabilities and how expectiles can be used to estimate the distribution function. The resulting fitted distribution function estimator can be inverted leading to quantile estimates. We run a simulation study to investigate and compare the efficiency of the expectile based estimator.

5.1 Introduction

During the estimation of population parameters usually the mean is used as measure of choice. In some cases the median as centrality parameter is preferred. A generalization of the median is given by quantiles. Quantile estimation and quantile regression has seen a number of new developments in the last years with Koenker (2005) as central reference. The principle idea is thereby to estimate an inverted cumulative distribution function, generally called the quantile function $q_\alpha = F^{-1}(\alpha)$ for $\alpha \in (0,1)$, where the 0.5 quantile $q_{0.5}$ as median plays a central role. For survey data tracing from an unequal probability sample with known probabilities of inclusion Kuk (1988) shows how to estimate quantiles taking the inclusion probabilities into account. The central idea is thereby to estimate a distribution function of the variable of interest and invert this in a second step to obtain the quantile function. Chambers and Dunstan (1986) propose a model-based estimator of the distribution function which is extended in Rao, Kovar, and Mantel (1990) towards unequal probability sampling. Bayesian approaches in this direction have recently been proposed in Chen, Elliott, and Little (2010) and Chen, Elliott, and Little (2012).

Quantile estimation results by minimizing an L_1 loss function as demonstrated in Koenker (2005). If the L_1 loss is replaced by the L_2 loss function one obtains so called expectiles as introduced in Aigner, Amemiya, and Poirier (1976) or Newey and Powell (1987). For $\alpha \in (0,1)$ this leads to the expectile function m_α which, like the quantile function q_α, uniquely defines the cumulative distribution function $F(y)$. Expectile estimation has recently gained some interest, see e.g. Schnabel and Eilers (2009b), Pratesi, Ranalli, and Salvati (2009), Sobotka and Kneib (2012) or Guo and Härdle (2013). However since expectiles lack an interpretation as simple as quantiles their acceptance and usage in statistics is less developed than quantiles, see Kneib (2013). Quantiles and expectiles are connected in that a unique and invertible transformation function $h : [0,1] \rightarrow [0,1]$ exists so that $m_{h(\alpha)} = q_\alpha$, see Yao and Tong (1996) or De Rossi and Harvey (2009). This connection can be used to estimate quantiles from a set of fitted expectiles. This idea has been used in Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014) and the authors demonstrate in simulations that the resulting quantiles can be more efficient than empirical quantiles, even if a smoothing step is applied to the latter (see Jones, 1992). In this note we extend these findings and demonstrate how expectiles can be estimated for unequal probability samples and how to obtain a fitted distribution function from fitted expectiles.

5.2 Quantile Estimation

We consider a finite population with N elements and continuous survey variable $Y_1, \ldots, Y_N$. We are interested in quantiles of the cumulative distribution function $F(y) = \sum_{i=1}^{N} 1\{Y_i \leq y\}/N$ and define with

$$q_\alpha = \inf\{\arg\min_q \sum_{i=1}^{N} w_\alpha(Y_i - q) \mid Y_i - q \mid\} \tag{5.1}$$

the quantile function of Y (see Koenker, 2005), where

$$w_\alpha(\epsilon) = \begin{cases} \alpha & \text{for } \epsilon > 0 \\ 1 - \alpha & \text{for } \epsilon \leq 0. \end{cases}$$

Apparently the "inf" argument in (5.1) is required in finite populations to guarantee a unique functional definition of $F(y)$. We draw a sample from the population with known inclusion probabilities π_i, $i = 1, \ldots, N$ and assume that π_i is positively correlated with Y_i. Denoting with $y_1, \ldots, y_n$ the resulting sample we estimate the quantile function by replacing (5.1) through its weighted sample version

$$\hat{q}_{N,\alpha} = \inf\{\arg\min_q \sum_{j=1}^{n} \frac{1}{\pi_j} w_{\alpha,j} |y_i - q|\} \tag{5.2}$$

with $w_{\alpha,j} = w_\alpha(y_j - q)$ as defined above. It is easy to see that the sum in (5.2) is an unbiased estimate for the sum in q_α given in (5.1). Nonetheless, because we take the "arg min" we do not obtain that $\hat{q}_{N,\alpha}$ is unbiased for q_α. In fact, it is easily shown that $\hat{q}_{N,\alpha}$ is the inverse of the normed weighted cumulative distribution function

$$\hat{F}_N(y) := \frac{\sum_{j=1}^{n} 1\{y_j \leq y\}/\pi_j}{\sum_{j=1}^{n} 1/\pi_j}$$

using the same notation as in Kuk (1988). Note that $\hat{F}_N(y)$ is not a Horvitz-Thompson estimate and as a consequence $\hat{q}_{N,\alpha}$ it is not unbiased. Nonetheless, $\hat{F}_N(y)$ is a proper distribution function, and hence it can be considered as normalized version of the Lahiri or Horvitz-Thompson estimator of the distribution function (see Lahiri, 1951) which is denoted by

$$\hat{F}_L(y) := \frac{1}{N}\sum_{j=1}^{n} 1/\pi_j 1\{y_j \leq y\}.$$

Kuk (1988) proposes to replace $\hat{F}_L(\cdot)$ with alternative estimates of the distribution function: Instead of estimating the distribution function itself he suggests to estimate the complementary proportion $\hat{S}_R(y)$ which then leads to the estimate $\hat{F}_R(y)$ defined through

$$\hat{F}_R(y) = 1 - \hat{S}_R(y) = 1 - \frac{1}{N}\sum_{j=1}^{n} 1/\pi_j 1\{y_j > y\}.$$

Resulting directly from these definitions we can express $\hat{F}_R(\cdot)$ in terms of $\hat{F}_N(\cdot)$ through

$$\hat{F}_R = 1 - \frac{1}{N}\sum_{j=1}^{n} 1/\pi_j + \hat{F}_L \quad \text{and} \quad \hat{F}_L = \frac{\sum_{j=1}^{n} 1/\pi_j}{N}\hat{F}_N. \tag{5.3}$$

Kuk (1988) shows that, under sampling with unequal probabilities, estimation of the median derived from $\hat{F}_R$ outperforms median estimates derived from $\hat{F}_N$ and $\hat{F}_L$ in terms of mean squared estimation error. Note that in the case of a simple random sample where $\pi_j = \pi = n/N$ the estimators $\hat{F}_N$, $\hat{F}_L$ and $\hat{F}_R$ coincide.

5.3 Expectile Estimation

An alternative to quantiles are expectiles. The expectile function m_α is thereby defined by replacing the L_1 loss in (5.1) by the L_2 loss leading to

$$m_\alpha = \arg\min_m \{\sum_{i=1}^{N} w_\alpha(Y_i - m)(Y_i - m)^2\}. \tag{5.4}$$

Note that m_α is continuous in α even for finite populations. Moreover $m_{0.5}$ equals the mean value $\bar{Y} = \sum_{i=1}^{N} Y_i/N$. Using the sample $y_1, \ldots, y_n$ with inclusion probabilities $\pi_1, \ldots, \pi_n$ we can estimate m_α by replacing the sum in (5.2) by its sample version, i.e.

$$\hat{m}_\alpha = \arg\min_m \{\sum_{j=1}^{n} \frac{1}{\pi_j} w_{\alpha,j}(y_j - m)^2\}$$

with $w_{\alpha,j}$ as defined above. It is easy to see that the sum in $\hat{m}_\alpha$ is an unbiased estimate for the sum in m_α. The estimate itself is however not unbiased like for the quantile function from Section 5.2.

5.4 From Expectiles to the Distribution Function

Note that both, the quantile function q_α and the expectile function m_α uniquely define a distribution function $F(.)$. While q_α is just the inversion of $F(.)$ the relation between m_α and $F(.)$ is more complicated. Following Schnabel and Eilers (2009b) and Yao and Tong (1996) we have the relation

$$m_\alpha = \frac{(1-\alpha)(m_\alpha) + \alpha\Big(m_{0.5} - G(m_\alpha\Big)}{(1-\alpha)F(m_\alpha) + \alpha\Big(1 - F(m_\alpha\Big)} \tag{5.5}$$

where $G(m)$ is the moment function defined through $G(m) = \sum_{i=1}^{N} Y_i 1\{Y_i \leq m\}/N$. Formula (5.5) relates function m_α to the distribution function $F(.)$. The idea is now to solve (5.5) for $F(.)$, that is to express the distribution $F(.)$ in terms of the expectile function. Apparently, this is not possible in analytic form but we may calculate this numerically. To do so, we evaluate the fitted function $\hat{m}_\alpha$ at a dense set of values $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_L < 1$ and denote the fitted values as $\hat{m}_l = \hat{m}_{\alpha_l}$. We also define left and right bounds through $\hat{m}_0 = \hat{m}_1 - c_0$ and $\hat{m}_{L+1} = \hat{m}_L + c_{L+1}$, where c_0 and c_L are some constants to be defined by the user. For instance one may set $c_0 = \hat{m}_2 - \hat{m}_1$ and $c_{L+1} = \hat{m}_L - \hat{m}_{L-1}$. By doing so we derive fitted values for the cumulative distribution function $F(.)$ at $\hat{m}_l$ which we write as $\hat{F}_l := \hat{F}(\hat{m}_l) = \sum_{j=1}^{l} \hat{\zeta}_l$ for non-negative steps $\hat{\zeta}_j \geq 0, j = 1, \ldots, L$ with $\sum_{j=1}^{L} \hat{\zeta} \leq 1$. We define $\hat{\zeta}_{L+1} = 1 - \sum_{l=1}^{L} \hat{\zeta}_l$ to make $\hat{F}(.)$ a distribution function. Assuming a uniform distribution between the dense supporting points $\hat{m}_l$ we may express the moment function $G(.)$ as

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

where $\hat{c}_j = (\hat{m}_j - \hat{m}_{j-1})/2$ and $\hat{G}_{L+1} = \hat{M}(0.5) = \sum_{j=1}^{n} (y_j/\pi_j) / \sum_{j=1}^{n} (1/\pi_j)$. With the steps $\hat{\zeta}_l, l = 1, \ldots, L$ we can now re-express (5.5) as

$$\hat{m}_l = \frac{(1-\alpha)\sum_{j=1}^{l} \hat{c}_j \hat{\zeta}_j + \alpha(\hat{m}_{0.5} - \sum_{j=1}^{l} \hat{c}_j \hat{\zeta}_j)}{(1-\alpha)\sum_{j=1}^{l} \hat{\zeta}_j + \alpha(1 - \sum_{j=1}^{l} \hat{\zeta}_j)}, \qquad l = 1, \ldots, L,$$

which is then be solved for $\hat{\zeta}_1, \ldots, \hat{\zeta}_L$. This is a numerical exercise which is conceptually rather straightforward. Details can be found in Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014). Once we have calculated $\hat{\zeta}_1, \ldots, \hat{\zeta}_l$ we have an estimate for the cumulative distribution function which is denoted as $\hat{F}_N^M(y) = \sum_{l:\hat{m}_l<y} \hat{\zeta}_l$. We may also invert $\hat{F}_N^M(.)$ which leads to a fitted quantile function which we denote with $\hat{q}_{N,\alpha}^M$.

As simulations and derivations in Kuk (1988) show, $\hat{F}_R(.)$ is more efficient than $\hat{F}_N(.)$ which led to the definition of $\hat{F}_R(.)$ given in (5.3). We take this relationship and apply it to $\hat{F}_N^M(.)$ which yields the estimator

$$\hat{F}_R^M := 1 - \frac{1}{N}\sum_{j=1}^{n} 1/\pi_j + \frac{\sum_{j=1}^{n} 1/\pi_j}{N}\hat{F}_N^M.$$

In the next section we will compare the quantiles calculated from the expectile based estimator $\hat{F}_R^M$ with quantiles calculated from $\hat{F}_R$. Note that neither $\hat{F}_R^M$ nor $\hat{F}_R$ are proper distribution functions since they are not normed to take values between 0 and 1.

5.5 Simulations

We run a small simulation study to show the performance of the expectile based estimates. We look at the two data sets also used in Kuk (1988). The first data set (Dwellings) contains the two variables X, the number of dwelling units, and Y, the number of rented units, which are highly correlated (with a correlation of 0.97), see Kish (1965). The second data set (Villages) includes information on the population (X) and on the number of workers in household industry (Y) for 128 villages in India, see Murthy (1967). In the second data set the correlation between Y and X is 0.54. In order to compare our simulation results with the results of Kuk (1988) we choose a same sample size of $n = 30$ (from a total population of $N = 270$ for the Dwellings data and $N = 128$ for the Villages data).

We compare quantiles defined by inversion of $\hat{F}_R$ with quantiles defined by inversion of $\hat{F}_R^M$. In Table 5.1 we give the root mean squared error (RMSE) and the relative efficiency for specified quantiles. We can see that the median and for the Dwelling data also upper quantiles derived from expectiles yield increased efficiency.

To obtain more insight we run a simulation scenario which involves a larger sample size of $n = 100$ selected from a population of size $N = 1000$ and $N = 10000$. We draw Y and X from a bivariate log standard normal distribution with expectation $\mu = 0$ and standard deviation $\sigma = 1$. Variables Y and X are drawn such that the correlation between the variables is equal to 0.9. We again calculate the root mean squared error for a range of α

	α	quantiles $\sqrt{\mathrm{MSE}(\hat{q}_{R,\alpha})}$	quantiles from expectiles $\sqrt{\mathrm{MSE}(\hat{q}^M_{R,\alpha})}$	relative efficiency $\frac{\sqrt{\mathrm{MSE}(\hat{q}^M_{R,\alpha})}}{\sqrt{\mathrm{MSE}(\hat{q}_{R,\alpha})}}$
Dwellings	0.1	2.57	2.76	1.07
	0.25	1.77	1.97	1.11
	0.5	2.45	2.35	0.96
	0.75	3.15	2.91	0.92
	0.9	4.20	3.43	0.82
Villages	0.1	5.52	6.65	1.21
	0.25	11.41	10.31	0.90
	0.5	12.29	11.69	0.95
	0.75	16.24	15.41	0.95
	0.9	13.31	18.34	1.38

Table 5.1: Comparison of mean squared error on a basis of 500 replications.

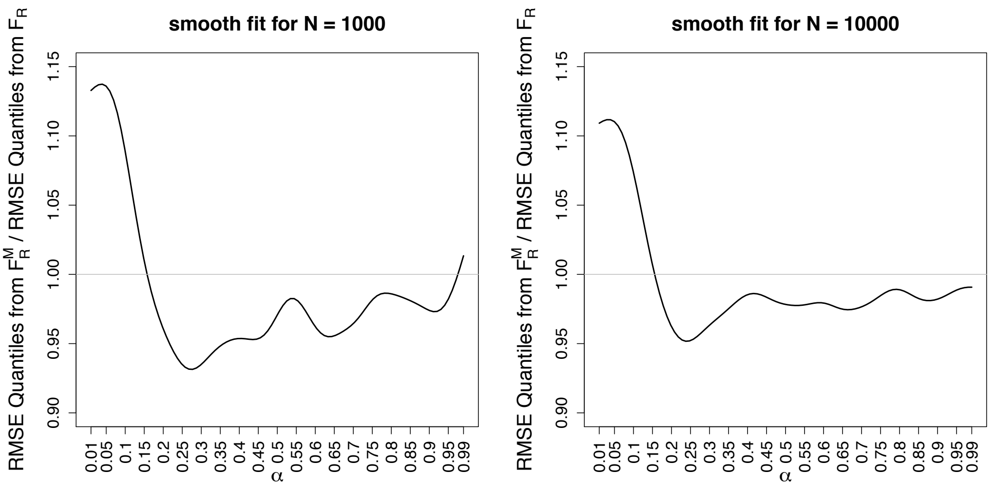


Figure 5.1: Relative Root Mean Squared Error (RMSE) of quantiles and quantiles from expectiles for the PPS design calculated from 500 repetitions (left: $N = 1000$, right: $N = 10000$).

values and show the relative efficiency of the expectile based approach in Figure 5.1. For better visual presentation we show a smoothed version of the relative efficiency. We notice a reduction in the root mean squared error for both cases $N = 1000$ and $N = 10000$. We may conclude that the expectiles can easily be fitted in unequal probability sampling and the relation between expectiles and the distribution function can be used numerically to calculate quantiles with partly increased efficiency.

5.6 Discussion

In Section 5.4 we extended the toolbox of expectiles to the estimation of distribution functions in the framework of unequal probability sampling. We defined expectiles for unequal probability samples. When comparing quantiles based on $\hat{F}_R$ with quantiles based on the expectile based estimator $\hat{F}_R^M$, we saw that the newly gained estimator can compete with existing methods. The calculation of empirical expectiles is implemented in the open source software R (see R Core Team 2014) and can be found in the R package expectreg by Sobotka, Schnabel, and Schulze Waltrup (2013). The calculation of the expectile based distribution function estimator $\hat{F}_N^M$ is also part of the R package expectreg.

Chapter 6

expectreg: An R Package for Expectile Regression

This chapter bases on a manuscript developed in joint work with Fabian Sobotka, Sabine Schnabel, Göran Kauermann and Thomas Kneib. A draft version of the manuscript is already published as part of the thesis *Semiparametric Expectile Regression* by Fabian Sobotka, the leading author of the paper. Most of the sections were written by Fabian Sobotka. The subsection on boosting was written by Fabian Sobotka and Thomas Kneib and the subsection on expectile sheets was written by Sabine Schnabel. Linda Schulze Waltrup contributed and wrote the subsections "Quadratic programming" and "Quadratic programming CDF". All authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript.

Abstract

Expectiles are a flexible least squares generalization of the mean similar as quantiles are an asymmetric generalization of the median. They are becoming more and more popular especially in semiparametric additive regression settings. A variety of estimation procedures including boosting and quadratic programming is available and they are combined in the R package `expectreg`. It also includes confidence intervals for the basic least asymmetrically weighted squares estimates. Latest methods also tackle the problem of crossing expectiles which is well known from quantile regression. Additionally, methods to compute quantiles from expectiles are described and included in the package. All functionalities are illustrated in examples using textbook data sets as well as those coming with `expectreg`.

6.1 Introduction

In the past years, methods that aim to gather more information than a mean regression have grown in popularity. A semiparametric method to fit more than the median is presented by quantile regression (Koenker and Bassett, 1978). In the R package `quantreg` (Koenker 2013b) single regression quantiles are estimated with the use of linear programming techniques. The estimation of a set of non-crossing quantiles was also proposed by Bollaerts, Eilers, and Aerts (2006) and Bondell, Reich, and Wang (2010), but code for the procedure is only partly available for download and not in package form. A different way to estimate quantiles was introduced by Efron (1991). It uses a least asymmetrically weighted squares (LAWS) approach for the estimation of linear regression curves that are then matched to quantiles. The result of a LAWS estimation is called expectile. Before the introduction of our package `expectreg` (Sobotka, Schnabel, and Schulze Waltrup, 2013) it was only possible to estimate expectiles manually by iterating a weighted mean regression, for example with the function `lm` from R or with the package `mgcv` (Wood, 2013). The package `SemiPar` (Wand, 2013) allows for semiparametric mixed models, but not for weights. The package `VGAM` (Yee, 2012) includes an option for expectile regression in the presence of the assumption of a distribution for the response. However, this leads to parallel expectile curves. Thus heteroscedasticity in the data cannot be captured and the gain of information in comparison to a mean regression is limited.

In contrast, generalized additive models for location, scale and shape (GAMLSS) follow a similar goal with different methods. GAMLSS were developed by Rigby and Stasinopoulos (2005) and estimate the mean together with additional parameters of the response distribution using backfitting. Alternatively, Mayr, Fenske, Hofner, Kneib, and Schmid (2012) proposed the estimation of GAMLSS with model based boosting. For both approaches, R packages are available: `gamlss` (Stasinopoulos and Rigby, 2013) and `gamboostLSS` (Hofner, Mayr, Fenske, and Schmid, 2011).

Our package not only contains methods to deal with univariate expectiles, but also includes e.g. all regression methods described and introduced in Schnabel and Eilers (2009b), Schnabel and Eilers (2014), Sobotka and Kneib (2012), Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013) and also Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014). Hence, `expectreg` offers regression estimates based on least squares and boosting, location-scale models and more refined techniques to overcome crossing curves.

The package is available from the Comprehensive R Archive Network at
`http://CRAN.R-project.org/package=expectreg`.

The rest of the paper is organized as follows. In Section 6.2 we present definitions of expectiles in comparison to quantiles and the corresponding functions from the package. Functionality for computation of theoretical expectiles as well as the estimation of sample expectiles is provided. Afterwards we introduce expectile regression methods in Section 6.3. We present different methods for the estimation of expectile curves and the computational issues arrising. The possibilities to compose distribution estimates from expectiles, both univariate and multivariate, are discussed in Section 6.4. Next we extend our regression methods to additive semiparametric models and shape constraint effects in Section 6.5. Section 6.6 shows the package in action for example data sets that are also included in `expectreg`. Finally, Section 6.7 concludes our findings by discussing general issues of expectile and quantile regression. We also supply an outlook into future extensions of the package.

6.2 Univariate expectiles

Let Y be a continuous random variable with density $f(y)$. Then the α-quantile q_α, $\alpha \in (0,1)$ is defined implicitly by the equation

$$\alpha = \mathbb{P}(Y \leq q_\alpha) = \frac{\int_{-\infty}^{q_\alpha} f(y)dy}{\int_{-\infty}^{\infty} f(y)dy} = \int_{-\infty}^{q_\alpha} f(y)dy.$$

This implicit definition characterizes the quantile as the partial integral of the density. A dense set of quantiles characterizes the corresponding distribution in a mostly sufficient way. The same holds for expectiles. The α-expectile m_α can be obtained by replacing the partially integrated density with the partial moment function G, yielding

$$\alpha = \frac{\int_{-\infty}^{m_\alpha} |y - m_\alpha| f(y)dy}{\int_{-\infty}^{\infty} |y - m_\alpha| f(y)dy} = \frac{G(m_\alpha) - m_\alpha F(m_\alpha)}{2(G(m_\alpha) - m_\alpha F(m_\alpha)) + (m_\alpha - m_{0.5})} \tag{6.1}$$

where F is the cdf, $G(m) = \int\limits_{-\infty}^{m} yf(y)dy$ and $G(\infty) = m_{0.5}$ is the expectation of Y.

For every given distribution with cumulative distribution function F and finite expectation we can compute all theoretical α-expectiles. Unfortunately the solution cannot be determined explicitly as quantiles from the inverse distribution function. By using Equation (6.1), the theoretical expectiles can be computed via a numerical procedure in R like `nlm` or `optimize`. In addition, the integrals of partial moments of a set of commonly used distributions have an analytical solution as shown by Winkler, Roodman, and Britney (1972). Exploiting these results and the numerical optimizations, we were able to implement R functions calculating expectiles for the most common distributions in the package. The R functions minimize

$$h(y) = \left| \frac{G(y) - yF(y)}{2(G(y) - yF(y)) + (y - m_{0.5})} - \alpha \right|$$

using `nlm`, resulting in $m_\alpha = \arg\min_z h(z)$. The functions follow the regular naming convention of distributions in R consisting of a letter (`p,d,q,r`) determining the use of the function and an abbreviation of the name of the distribution. For example, `qnorm` computes the quantiles (`q`) of the normal (`norm`) distribution. Our expectile functions extend this set and are named accordingly starting with the letter "`e`", resulting in `enorm` for the expectiles of a normal distribution. An overview of the functions for distribution-expectiles is given in Table 6.1.

Furthermore our package contains a distribution for which expectiles and quantiles coincide. The distribution function is defined as

$$F_{emq}(y; m, s) = \frac{1}{2}\left(1 + sign\,(y - m)\sqrt{1 - \frac{2}{2 + \left(\frac{y-m}{s}\right)^2}}\right)$$

with density

$$f_{emq}(y; m, s) = \frac{1}{s}\left(\frac{1}{2 + \left(\frac{y-m}{s}\right)^2}\right)^{\frac{3}{2}}$$

for $s > 0$ and $y, m \in \mathbb{R}$. The property of $e_{emq}(\alpha; m, s) = q_{emq}(\alpha; m, s)$ holds for all possible expectations m, scalings s and all asymmetries α. Therefore we named it the "expectiles meet quantiles" (emq) distribution. A special case of the family with parameters $m = 0$ and $s = \sqrt{2}$ was mentioned by Koenker (1993). For the canonical parameters $m = 0$ and $s = 1$ we get a student-t distribution with two degrees of freedom. This family of distributions does not have finite second moments regardless of the value of the scaling parameter s. The expectation for a random variable $Y \sim \texttt{EMQ}(m, s)$ is $E(Y) = m$. The software provides a complete set of functions for this distribution in the common naming conventions as mentioned above. For completeness the package contains also the numerical determination of expectiles in the function `eemq`.

In order to estimate expectiles from a given sample, we exploit a similarity between the calculation of expectiles and quantiles. For given i.i.d. observations $y_1, \ldots, y_n$ the α-quantile

Distribution	Expectile function	Parameters
normal	`enorm`	m, sd
student t	`et`	df
χ^2	`echisq`	df
gamma	`egamma`	shape, rate, scale
exponential	`eexp`	rate
beta	`ebeta`	a,b
uniform	`eunif`	min, max
lognormal	`elnorm`	meanlog, sdlog
emq	`eemq` / `qemq`	m, s

Table 6.1: List of theoretical expectile functions implemented in the package including the distribution parameters.

can be determined by minimizing

$$S = \sum_{i=1}^{n} w_\alpha(y_i, q_\alpha)|y_i - q_\alpha| \tag{6.2}$$

with weights

$$w_\alpha(y_i, q_\alpha) = \begin{cases} \alpha & \text{if } y_i > q_\alpha \\ 1-\alpha & \text{if } y_i \leq q_\alpha \end{cases}, \tag{6.3}$$

where q_α is the resulting α-quantile of the observations. The same can be achieved for expectiles using a least squares minimization criterion. Least asymmetrically weighted squares (LAWS) is a weighted generalisation of ordinary least squares (OLS) estimation. LAWS minimizes

$$S = \sum_{i=1}^{n} w_\alpha(y_i, m_\alpha)(y_i - m_\alpha)^2, \tag{6.4}$$

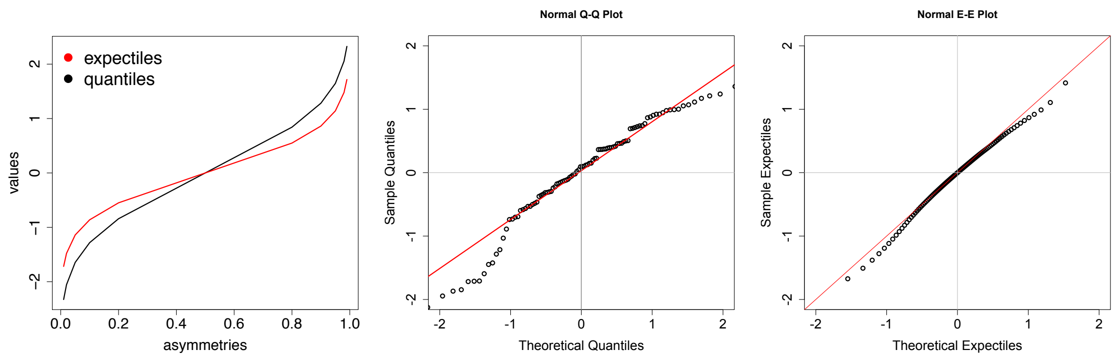
with weight function $w_\alpha(y_i, m_\alpha)$ according to Equation (6.3), observations y_i and population expectile m_α for different values of an asymmetry parameter $\alpha \in (0,1)$. The expectile is fitted by alternating between a weighted least squares fit and recomputing weights until convergence of the weights. Equal weights of $w_\alpha(y_i, m_\alpha) = 0.5$ are obviously a convenient starting point. They also make up the special case of OLS.

In R we are able to use the expectile functions just like quantiles. Because of their implicit definition, the calculation of theoretical distribution expectiles is numerically more challenging and therefore not as fast as the computation of quantiles, but the results are equally useful. The estimation of expectiles, on the other hand, is still a fast procedure since it bases on a least squares fit. For univariate samples, we provide the function `expectile` that returns a vector of fitted expectiles similar to the function `quantile` that comes with R. The asymmetries to be computed are accordingly named `probs` for convenience although the probability statement that is connected to the computed expectiles is different. In the same manner, the functions `eeplot` and `eenorm` are pendants to the existing quantile functions. The latter plots the results of the function `enorm` against those from `expectile` for α in steps of 0.01.

With the following example, we can show that a set of expectiles is equally appropriate to get an impression of the distribution as a similar set of quantiles. We show expectiles and quantiles from a standard normal distribution for selected values of α between 0.01 and 0.99.

```
R> alpha = c(0.01,0.02,0.05,0.1,0.2,0.5,0.8,0.9,0.95,0.98,0.99)
R> plot(alpha,qnorm(alpha))
R> lines(alpha,enorm(alpha),col="red")
R> set.seed(126)
R> y = rnorm(1000)
R> quantile(y,probs=alpha)
R> expectile(y,probs=alpha)
R> qqnorm(y)
R> eenorm(y)
```

Both the quantiles and the expectiles from the standard normal distribution are displayed in Figure 6.1 and show the symmetry of the distribution, its median or mean and the variability. We can also see the satisfying quality of the estimated expectiles from a standard normally distributed sample of size 1000. When comparing the estimated expectiles and quantiles with their true values, as for example in Figure 6.1, we can see that the E-E plot shows a stronger support for the normal distribution than the Q-Q plot. For the normal distribution it is known that expectiles have a higher asymptotic relative efficiency than quantiles (Abdous and Remillard, 1995), hence the result is not surprising. The results of Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014) also show that, if the variance exists, univariate expectiles can be estimated at least as efficient as quantiles. The paper by Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014) also introduces a novel method to obtain quantiles from expectiles. These findings further motivate the use of expectiles, also in regression scenarios.



(a) `enorm(alpha); qnorm(alpha)` (b) `qqnorm(y)` (c) `eenorm(y)`

Figure 6.1: Visualisation of quantiles and expectiles from a standard normal distribution. Graph (a) displays the theoretical quantiles and expectiles. Graphs (b) and (c) show a Q-Q plot and an E-E plot in comparison for a sample of 1000 random numbers drawn from a standard normal distribution.

6.3 Expectile regression

In the past years, several methods for the estimation of expectiles in semiparametric regression settings have been developed. In the least asymmetrically weighted squares approach, expectiles are estimated separately for each asymmetry. Since that can lead to crossing curves, location-scale models but also more flexible "sheets" have been proposed that lead to non-crossing results. The estimation can rely on different computational concepts as least squares, boosting and quadratic programming. All tools are introduced in this section.

6.3.1 Penalized estimation concepts

LAWS

Let us first consider a simple parametric model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_\alpha + \boldsymbol{\epsilon}_\alpha$$

with response $\boldsymbol{y}$, covariate matrix $\boldsymbol{X}$ and expectile $m_\alpha = \boldsymbol{X}\boldsymbol{\beta}_\alpha$ and errors $\boldsymbol{\epsilon}_\alpha$. In an expectile regression setting, the estimation of the regression coefficients that minimize Equation (6.4) can be accomplished by iteratively reweighted least squares updates

$$\hat{\boldsymbol{\beta}}_\alpha^{[b]} = (\boldsymbol{X}'\boldsymbol{W}_\alpha^{[b-1]}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}_\alpha^{[b-1]}\boldsymbol{y} \tag{6.5}$$

where $\hat{\boldsymbol{\beta}}_\alpha^{[b]}$ is the regression coefficient vector in the bth iteration corresponding to the model. Note that estimation has to be iterated since the weights gathered in weight matrix $\boldsymbol{W}_\alpha^{[b]} = \mathrm{diag}(w_\alpha(y_1, \boldsymbol{X}\hat{\boldsymbol{\beta}}_\alpha^{[b]}), \dots, w_\alpha(y_n, \boldsymbol{X}\hat{\boldsymbol{\beta}}_\alpha^{[b]}))$ also depend on the current estimates.

Next, we look at a more flexible nonlinear model

$$\boldsymbol{y} = f_\alpha(\boldsymbol{x}) + \boldsymbol{\epsilon}_\alpha\,.$$

For the expectile curve f_α several choices for the functional form are possible. The original proposal by Newey and Powell (1987) favored a linear model. We suggest a more flexible functional form for the expectile curve. Schnabel and Eilers (2009b) proposed to model expectile curves with P-splines.

The basic idea of spline smoothing is to approximate the unknown function $f(x)$ by a polynomial spline of degree l

$$f(x) = \sum_{k=1}^{K} u_k B_k^{(l)}(x)$$

where $B_k^{(l)}(x)$ are B-spline basis functions, u_k are the corresponding amplitudes, and K denotes the dimensionality of the basis. We can summarize this into a design matrix $\boldsymbol{B}$ using the basis function evaluations while the amplitudes are collected in the vector of regression coefficients $\boldsymbol{u}$. The form of the polynomial spline approximation crucially depends on the number and location of the knots. To overcome this difficulty, penalized splines use a set of K equidistant knots in combination with a smoothness penalty augmented to the fitting criterion. A popular choice inspired by smoothing splines are penalties based on integrated squared derivatives as in Wood (2006). However, we prefer to work with the simpler approximation of Eilers and Marx (1996) based on differences of adjacent coefficients. Let $\boldsymbol{D}$ denote a difference matrix of order r, then the penalty matrix $\boldsymbol{P} = \lambda\boldsymbol{D}'\boldsymbol{D} = \lambda\boldsymbol{K}$ yields a penalty composed of a scalar smoothing parameter λ and squared r-th order differences in the sequence of basis coefficients, i.e., $\boldsymbol{u}'\boldsymbol{P}\boldsymbol{u} = \sum_{k=r+1}^{K}(\Delta_r(u_k))^2$ where Δ_r is the r-th order difference operator. Common choices are $K = 20$ and $r = 2$. The estimation of the regression coefficients is now the iteration between calculating

$$\hat{\boldsymbol{u}}_\alpha^{[b]} = (\boldsymbol{B}'\boldsymbol{W}_\alpha^{[b-1]}\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{W}_\alpha^{[b-1]}\boldsymbol{y}$$

and recomputing the weights as described before.

Restricted expectiles

In theory, expectiles are as monotonous along α as quantiles, but in regression scenarios we sometimes encounter crossings of quantile or expectile curves due to sampling variation and/or sparse data. As a consequence, He (1997) proposed restricted regression quantiles to avoid the crossing of quantile curves. His model for computing nonparametric conditional quantile functions takes the following location-scale form

$$q(x, \alpha) = t(x) + s(x)c_\alpha. \tag{6.6}$$

He (1997) introduced a three-step procedure where first the conditional median function $t(x)$ is determined and then in a second step the smooth non-negative amplitude function $s(x)$ is estimated. The third step consists of the stepwise calculation of the "asymmetry factor" c_α for each α-quantile curve separately. The obvious advantage is that we only need monotonicity in c_α to obtain non-crossing quantile curves. In return, we give up a large amount of flexibility and are not able to capture all possible information in heteroscedastic scenarios. In quantile regression, however, crossing curves are very common and the estimated restricted quantiles might still deliver better results than the independent fits of quantiles.

Although we observe less crossings in the independent estimation of expectile curves, we still adapted the concept of restricted quantiles to expectile regression. The loss of flexibility might still be preferable in a scenario where a lot of crossing expectiles are observed. In the provided algorithm, the previously described three estimation steps are matched to estimate restricted expectiles. First the mean function $t(x)$ is estimated as a mean regression by the chosen computational method. Then the residuals from the first step are used to estimate $s(x)$ in the same way. Finally, c_α is estimated as regression coefficient in the expectile regression with response $y - t(x)$ and covariate $s(x)$. In this case we use the LAWS loss function as described in the previous section.

LAWS bundle

The expectile bundle –as introduced in Schnabel and Eilers (2013a)– has strong similarities to the restricted expectiles. It is based on the location-scale model defined in Equation (6.6) and essentially it also follows three estimation steps. The main difference lies in an additional iteration of the second and third step. In step 2, LAWS is used to fit the residual curve $s(x)$ optimally to all calculated expectiles. Then the asymmetry parameters c_α are recomputed and both steps are repeated until convergence. For the second step and the set of selected asymmetries $\alpha_1 < \cdots < \alpha_T$, we perform an expectile regression with response $\tilde{y} = ((y_1 - m(0.5, \boldsymbol{x}_1))^{(1)}, \ldots, (y_n - m(0.5, \boldsymbol{x}_n))^{(1)}, \ldots, (y_1 - m(0.5, \boldsymbol{x}_1))^{(T)}, \ldots, (y_n - m(0.5, \boldsymbol{x}_n))^{(T)})'$, weight matrix

$$\boldsymbol{W} = \text{diag}(w_{\alpha_1}(\tilde{y}_1, m_{\alpha_1}), \ldots, w_{\alpha_1}(\tilde{y}_n, m_{\alpha_1}), \ldots, w_{\alpha_T}(\tilde{y}_n, m_{\alpha_T}))$$

and the appropriate repetitions of design matrix $\tilde{\boldsymbol{B}} = (c_{\alpha_1}\boldsymbol{B}, \ldots, c_{\alpha_T}\boldsymbol{B})'$ and penalty matrix. We then obtain an estimate for the residual function $s(x)$ that is optimal for the selected set of asymmetries.

Expectile sheets

A more flexible approach to expectile smoothing are the so-called expectile sheets. Every set of expectile curves for a dense set of asymmetry parameters α forms a surface over the domain of (x, α). With an expectile sheet, this plane is fitted directly, thus forming a natural description of a comprehensive set of expectile curves. While in a simple LAWS model every α-expectile is fitted separately, all expectile curves are modeled simultaneously in an expectile sheet.

The sheet is the result of minimizing

$$S_{ES} = \sum_{i=1}^{n} \sum_{t=1}^{T} w_{\alpha_t}(y_i, m_{\alpha_t})(y_i - m(x_i, \alpha_t))^2 \tag{6.7}$$

with the weights $w_{\alpha_t}(y_i, m_{\alpha_t})$ as defined above and $m(x, \alpha)$ is the expectile sheet.

The expectile sheet itself can be formulated as a tensor product of two B-spline bases, one basis over x and one basis over α. Bivariate splines are described in detail in Section 6.5.1. When dealing with complex datasets and a dense set of asymmetries, computation of an expectile sheet can be costly. Therefore Schnabel and Eilers (2014) suggested to make use of fast array algorithms for multidimensional P-spline fitting (see Currie, Durbán, and Eilers, 2006 and Eilers, Currie, and Durbán, 2006). As a consequence the construction of the Kronecker product is avoided and this results in speeding up the calculations.

6.3.2 Computational methods and implementation

Least squares

Within the function `expectreg.ls` the R procedure `lsfit` is responsible for the estimation, using the response $(\boldsymbol{y}, \boldsymbol{0}_{(p+1)})'$, the design matrix $(\boldsymbol{B}', \sqrt{\lambda}\boldsymbol{P}')'$ and the weights $\boldsymbol{W}$. Then the weights are updated according to the resulting estimate. This process is iterated until the weights and therefore the estimates converge.

The function `rb` allows for building a penalized spline basis from a vector of covariate observations. Therefore by default 20 equidistant knots are chosen and the penalty matrix is created from second order differences. The evaluated spline basis is obtained from the package `splines` (R Development Core Team 2010). Optionally, the eigen decomposition of the penalty matrix is used to reparameterize the basis according to Fahrmeir, Kneib, and Lang (2004). We make use of the reparameterization to remove the intercept from the spline basis, since there is a central intercept in the fitting function `expectreg.ls`. Otherwise, setting the option `center=FALSE` can suppress the reparameterisation.

The package includes two algorithms to find the optimal value for the smoothing parameter λ. Both the Schall algorithm as well as asymmetric cross-validation are introduced in Schnabel and Eilers (2009b). The Schall algorithm (see Schall, 1991) is implemented as an iterative procedure that measures the numerical convergence of λ. The cross-validation on the other hand uses numerical minimization methods, in this case `nlm`, to find the minimum of the cross-validation score V_g^w with respect to the smoothing parameter. The score is defined as

$$V_g^w = \frac{n \sum_{i=1}^n w_\alpha(y_i, \boldsymbol{B}\hat{\boldsymbol{u}}_\alpha)(y_i - \boldsymbol{B}\hat{\boldsymbol{u}}_\alpha)^2}{[\mathrm{tr}(\mathbb{1} - \boldsymbol{H}_\alpha)]^2}$$

with the hat matrix

$$\boldsymbol{H}_\alpha = (\boldsymbol{W}_\alpha)^{1/2}\boldsymbol{B}(\boldsymbol{B}'\boldsymbol{W}_\alpha\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}'(\boldsymbol{W}_\alpha)^{1/2}. \tag{6.8}$$

As defined in Schnabel and Eilers (2009b), the score utilizes the trace of the hat matrix which is returned by the procedure `lsfit`. The R function `rb` is also able to create other types of bases, but this will be discussed in Section 6.5.

For the least squares estimation method with asymmetric weights, asymptotic properties are also available (see Sobotka, Kauermann, Schulze Waltrup, and Kneib, 2013):

$$\hat{\boldsymbol{u}}_\alpha \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{u}_\alpha, \mathrm{Var}(\hat{\boldsymbol{u}}_\alpha)\right)$$

with

$$\mathrm{Var}(\hat{\boldsymbol{u}}_\alpha) = (\boldsymbol{B}'\boldsymbol{W}_\alpha\boldsymbol{B} + \boldsymbol{\lambda}_\alpha\boldsymbol{P})^{-1}\left\{\boldsymbol{B}'\boldsymbol{W}_\alpha^2\boldsymbol{B}\frac{(\boldsymbol{y} - \boldsymbol{B}\hat{\boldsymbol{u}}_\alpha)^2}{\mathbb{1} - \mathrm{diag}(\boldsymbol{H}_\alpha)}\right\}(\boldsymbol{B}'\boldsymbol{W}_\alpha\boldsymbol{B} + \boldsymbol{\lambda}_\alpha\boldsymbol{P})^{-1}, \tag{6.9}$$

where $\boldsymbol{H}_\alpha$ is the generalized hat matrix as defined in Equation (6.8).

Similar to all regression functions in R, the model is handed to `expectreg.ls` in a formula object. In the following example from the data set "lidar" from the R package `SemiPar`, we perform a least squares expectile regression with `laws` estimate to assess the influence of the distance that laser light travels on the logarithm of the received amount of light. We use cross-validation for choosing the smoothing parameter. The estimation method is chosen with `estimate`. The parameter `expectiles` influences the asymmetries which are calculated. The default setting estimates expectiles for the set $\alpha \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98, 0.99\}$. Alternatively, a custom set of expectiles can be specified in a vector. Last, there is an additional option `density`. If selected, a very dense set of expectiles $\alpha \in \{0.01, 0.02, \ldots, 0.99\}$ is fitted which allows for the use of expectiles in density estimation. This will be described in Section 6.4. The covariance matrices are computed if the option `ci = TRUE` is selected.

The function returns an object of class `expectreg`. It contains all relevant information of the regression such as the formula, covariates and response as well as the results in form of the fitted values $\hat{\boldsymbol{y}}_\alpha$, smoothing parameters λ and the regression coefficients $\hat{\boldsymbol{u}}_\alpha$. Further details of the regression results can also be obtained in a similar form as from other regression objects, since a number of methods are available as `print`, `resid`, `coef` or `plot`. The latter produces the content of Figure 6.3. Additionally, there is a `predict` method that also accepts a data frame of new data points as long as the columns are named according to the original data. A special `summary` function returns parametric effects and their significances while `confint` computes confidence intervals based on the asymptotics for the whole fit or, if specified, a subset of covariates.

```
R> library("SemiPar")
R> data("lidar")
R> exp.l <- expectreg.ls(logratio ~ rb(range,"pspline"),
     data=lidar,smooth="acv",estimate="laws")
```

Quadratic Programming

In this subsection we will describe and illustrate the use of the function `expectreg.qp`. As `expectreg.ls`, function `expectreg.qp` allows for the estimation of a set of expectiles given covariate(s). Assuming an additive influence of the covariates on the response, the dependency may then be modeled linear or nonlinear. The estimation procedure within function `expectreg.qp` relies on the quadratic form of the estimation problem. The problem presented in Equation (6.4) can be seen as a quadratic programming problem which can be solved by the R function `solve.QP` implemented in package `quadprog` (see Turlach and Weingessel, 2013). The estimation process in general is – as described before – an iterative procedure. Weights are updated until convergence is reached. Instead of using `lsfit` we now use `solve.QP` because it offers an easy way to incorporate linear (inequality) constraints on the estimator. These constraints are constructed in such a way that non-crossing of the estimated expectiles is guaranteed. From the estimation concepts described in the previous section, only the sheets can be used with constraints.

We use a penalized spline basis of degree two in direction of covariate x and a linear penalized spline basis in direction of α. Non-crossing is ensured by adding linear constraints on the coefficients. We postulate that neighboring curves fulfill

$$m(\boldsymbol{x}, \alpha_t) \leq m(\boldsymbol{x}, \alpha_{t+1}) \quad \text{for } t = 1, \ldots, T-1.$$

Here, T denotes the dimensionality of the basis in direction of α and $\boldsymbol{x}$ is the vector of covariates. As we use function `solve.QP` for the estimation, linear constraints on the coefficients are easily incorporated.

Within the function `expectreg.qp` the linear spline basis in direction of asymmetry α is evaluated at knots $\alpha_1 < \cdots < \alpha_T$ which are specified by the user. The default setting in `expectreg` is $T = 11$ with the set of α specified in the previous sections since the default settings are the same in all estimation functions. Like with its sister `expectreg.ls`, in the function `expectreg.qp` bases are created with the function `rb`. The estimation of the smoothing parameter λ can be done automatically using the algorithm by Schall (1991). As the expectile $m(\boldsymbol{x}, \alpha)$ is constructed as a tensor product of B-splines (similar to 6.3.1), there is only one smoothing parameter for each smooth function which needs to be determined. Finally, the function `expectreg.qp` provides an object of class `expectreg` which allows access to all further postprocessing methods.

```
R> exp.qp <- expectreg.qp(logratio ~ rb(range,"pspline"),
      data=lidar)
```

Function `expectreg.qp` also allows for the inclusion of a random intercept. A random intercept is useful for modeling e.g. panel data as it adds a random shift for each individuum and therefore takes the special form of longitudinal data into account. To see how to estimate a random intercept with `expectreg.qp` consider the panel data from R package `plm` by Croissant and Millo (2008). The data consists of 19 observations from 1960 to 1978 and reports (among other variables) the logarithm of real per-capita income (`lincomep`) and the logarithm of real motor gasoline price (`lrpmg`) of 18 different countries on a yearly basis. We can now use function `expectreg.qp` to estimate the smooth functional relation between log per capita income and log gasoline price. To account for repeated measurements we also add a random intercept resulting in a random shift in log per-capita income for each country.

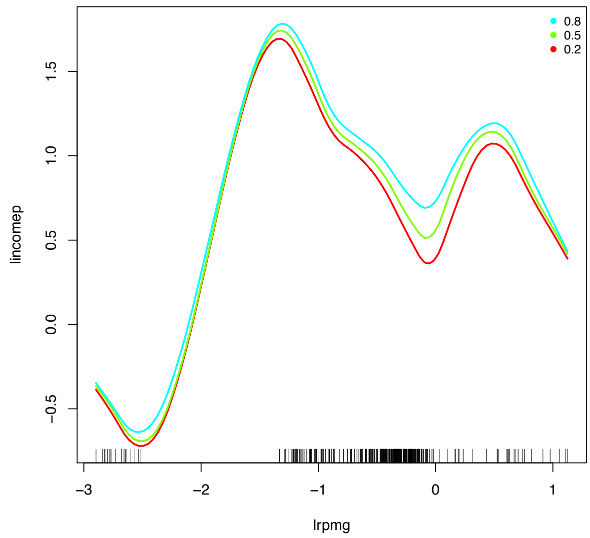


Figure 6.2: Expectile curves estimated for the Gasoline data.

```
R> library(plm)
R> data("Gasoline")
R> exp.qp.ri <- expectreg.qp(lincomep ~ rb(lrpmg, "pspline"),
      id = Gasoline$country, data = Gasoline, smooth = c("schall"),
      expectiles = c(0.2, 0.5, 0.8))
R> plot(exp.qp.ri)
```

The plot can be seen in Figure 6.2. In a next step we access the standard deviation of the residuals and the standard deviation of the random intercept.

```
R> sqrt(exp.qp.ri$sig2)
[1] 0.1799705
R> sqrt(exp.qp.ri$tau2)
[1] 0.3389326 0.4060940
```

The last element of `sqrt(exp.qp.ri$tau2)` gives the standard deviation of the random intercept.

Boosting

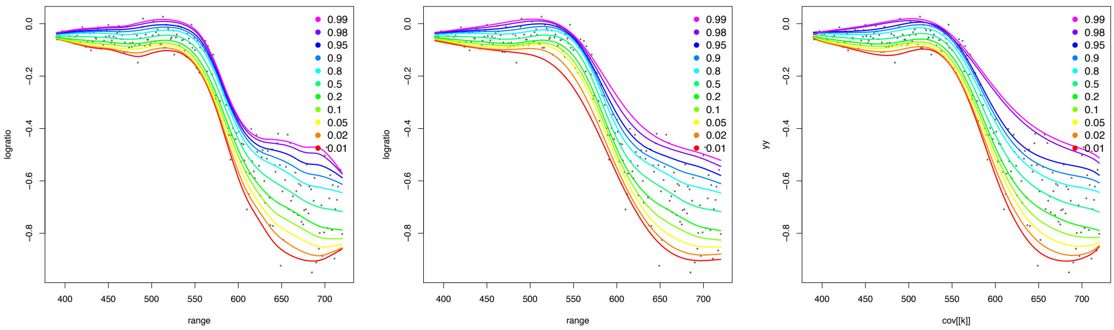
The package also contains a fitting method that provides an alternative to least squares fits. It combines model fitting with automatic variable selection and model choice. The function `expectreg.boost` performs componentwise functional gradient descent boosting to obtain the regression coefficients implemented via the package `mboost` (see Hothorn, Bühlmann, Kneib, Schmid, and Hofner, 2013). Bühlmann and Hothorn (2007) present a general introduction to boosting. The basic ingredient of a boosting algorithm are suitable base-learning procedures that are iteratively applied to the gradient vector $\boldsymbol{r}$ ("residuals") of the considered optimization criterion for a fixed number of iterations m_{stop}. In the expectile regression framework, a suitable class of base-learning procedures is given by penalized least squares estimates

$$\hat{\boldsymbol{g}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B} + \lambda\boldsymbol{P})^{-1}\boldsymbol{B}'\boldsymbol{r} = \boldsymbol{H}\boldsymbol{r},$$

characterized by the hat matrix $\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B}+\lambda\boldsymbol{P})^{-1}\boldsymbol{B}'$, see the next section for details on the specification of the design and penalty matrices. A comparable complexity of all base-learners avoiding estimation and selection bias is achieved when the smoothing parameters λ are chosen such that the degree of freedom $\text{df} = \text{tr}(2\boldsymbol{H} - \boldsymbol{H}'\boldsymbol{H}) = 1$ is comparable across all base-learners (Hofner, Hothorn, Kneib, and Schmid, 2011). Note that the chosen smoothing parameters do not assign a fixed amount of smoothness to the corresponding function estimates but only to the base-learner. Since the same base-learner may be chosen several times during the iterative fitting procedure, the ultimate function estimate can build up a much larger complexity. The optimal number of boosting iterations is usually determined via cross-validation techniques.

The implementation of boosting in our package completely relies on `mboost`, a package that comes with various methods for model-based boosting. A family object for expectile regression was introduced to the `mboost` package that holds information about the loss function (6.4) and the corresponding gradient. The boosting functions can be handled similar to all regression functions. An object of class `mboost` is returned with several methods available. For our purposes, we chose the function `gamboost` for generalized additive models. Further the method `cvrisk` allows to conduct a ten-fold cross validation to determine the optimal value m_{stop}. The `mboost` object can then be reduced to the first m_{stop} iterations.

The function `expectreg.boost` provided in our package works as a wrapper for the `mboost` functions. It ensures that expectile regression using boosting works with the same interface as the rest of the expectile functions. Additionally it makes sure that the result is open to the expectile methods like `plot` and `predict`. We use the set of base-learners



(a) `expectreg.ls` (b) `expectreg.qp` (c) `expectreg.boost`

Figure 6.3: Expectile curves estimated using `expectreg.ls`, `expectreg.boost` (both with LAWS estimate) and `expectreg.qp` for the lidar data from `SemiPar`.

from `mboost` for expectile boosting (e.g. `bbs` for P-splines). In Figure 6.3, the same example as before was used. Especially the results for the 0.98 and the 0.99-expectile indicate that smoothing works differently for boosting. So in general, m_{stop} will be chosen larger than optimal, but can be set individually for each expectile. We expect a larger number of iterations for extreme asymmetries while in the special case of a mean regression the optimal fit is achieved quite early.

A function `quant.boost` is also available. The difference between `quant.boost` and `expectreg.boost` is the loss function. In this case Equation (6.2) and the corresponding gradient are used. The interface and the available methods stay the same.

```
R> exp.boost <- expectreg.boost(logratio ~ bbs(range,"pspline"),
                                data=lidar,mstop=rep(500,11))
```

6.4 Distribution estimation from expectiles

Consider a random variable Y with distribution function $F(y) = P(Y \leq y)$. A way to look at the distribution of a random variable is via quantiles or expectiles. By definition quantiles and expectiles uniquely define the distribution of Y. That means, if we have a given set of expectiles, we should be able to construct the corresponding distribution function. The following methods are introduced in Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014) and Schnabel and Eilers (2013a).

6.4.1 Quadratic programming CDF

We will see how we can calculate the distribution function F from the expectile function m_α. First, we need a (dense) set of expectiles with corresponding asymmetry parameters $\alpha_t \in (0,1)$ where $t = 1, \dots, T$ and $\alpha_1 < \alpha_2 < \dots < \alpha_T$. In a next step we define $F_t := F(m_{\alpha_t})$ which needs to be determined from m_{α_t}. In particular we set

$$F_t = \sum_{j=1}^{t} \zeta_j, \quad t = 1, \dots, T$$

and calculate ζ_t from m_{α_t}. Note that ζ_t should be greater than zero for $t = 1, \dots, T$ and that $\sum \zeta_t$ should be smaller or equal one in order to guarantee that F is a distribution function. We start with specifying $F_0 \equiv 0$ and define the minimal expectile as m_0. (If the data from which the expectiles were calculated is still at hand, we can set m_0 as minimal observed value of the data. As this is not always the case we implemented the former version but use extreme expectiles as basis for the estimation: consider for example the 10^{-5}-expectile. This nearly corresponds to the minimal observed value.)

Remember that the partial moment is $G(m) = \int\limits_{-\infty}^{m} y f(y) dy$. We now replace density f through the approximated version $\tilde{f}$ which is defined as

$$\tilde{f}(y) = \begin{cases} \frac{\zeta_t}{m_{\alpha_t} - m_{\alpha_{t-1}}}, & \text{if } y \in [m_{\alpha_t}, m_{\alpha_{t-1}}), \\ 0, & \text{else} \end{cases}$$

and obtain

$$\tilde{G}(m_{\alpha_t}) := \int\limits_{-\infty}^{m_{\alpha_t}} y \tilde{f}(y) dy = \sum_{j=1}^{t} \frac{m_{\alpha_j} + m_{\alpha_{j-1}}}{2} \zeta_j.$$

In a next step we use the definition

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

and solve

$$g_t(\boldsymbol{\zeta}) := m_{\alpha_t} - \frac{(1-\alpha_t)G(m_{\alpha_t}) + \alpha(m_{0.5} - G(m_{\alpha_t})))}{(1-\alpha_t)F(m_{\alpha_t}) + \alpha(1 - F(m_{\alpha_t}))} = 0$$

subject to the constraint that $\zeta_t > 0$ for $t = 1, \dots, T$ numerically, where $\boldsymbol{\zeta}$ is defined as

$\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_T)$. We use the package quadprog (Turlach and Weingessel, 2013) to minimize $\frac{1}{2}\sum_{t=1}^{T} d_t^2(\boldsymbol{\zeta})$ within an iterative procedure (subject to the constraints mentioned above) and obtain $\tilde{\boldsymbol{\zeta}}$ as solution.

The procedure described above can be performed not only for population expectiles but offers as well the possibility to calculate the conditional cumulative distribution function (cdf) of a response y for a given value of covariate x. The function cdf.qp works on an object of class expectreg. The user can specify the covariate value where the cdf then is estimated. It is also possible to state a vector of probabilities and obtain the corresponding quantiles.

One of the data sets which are included in the package contains nine growth characteristics on 6848 dutch male children and is called dutchboys. The data set therefore contains measurements of the height and the age of the children along other variables as, for example, a centered BMI. We restrict our attention to the 200 youngest boys of the dutchboys-data. Suppose we want to estimate the cdf of the height of the children at a given age of 0.08 years. First we run an expectile regression in which we ensure the estimation of extreme expectiles. This improves the quality of the cdf which will be estimated afterwards by function cdf.qp. If we are not only interested in the cdf but also in quantiles of the distribution, we just have to specify the corresponding vector of probabilities within the function cdf.qp.

```
R> expectiles <- expectreg.ls(hgt ~ rb(age, "pspline"),
       data=dutchboys[1:200,],estimate = 'laws',
       expectiles=c(0.00001, seq(0.01,0.99, by=0.01),0.99999))
R> cdf <- cdf.qp(expectiles, 0.08, qout = c(0.05,0.2,0.5,0.8,0.95))
```

6.4.2 Bundle density

The second distribution estimation method specializes on calculating an overall density from an expectile regression based on a restricted or bundle estimate. In order to accomplish this distribution estimation, Equation (6.1) is rewritten to (see Godambe, 1991)

$$\int_{-\infty}^{\infty} f(y)\psi(y,\alpha)dy \quad = \quad 0 \qquad \forall \qquad (\alpha, m_\alpha)$$

with the asymmetric estimating function

$$\psi(y,\alpha) \quad = \quad \begin{cases} (1-\alpha)(y-m_\alpha) & y < m_\alpha \\ \alpha(y-m_\alpha) & y \geq m_\alpha \end{cases} .$$

Now we assume that m_{α_t} is known for a set of asymmetries $\alpha_1 < \alpha_2 < \ldots < \alpha_T$. This is an ill-posed problem for a bounded set of asymmetries and several densities could be constructed from the expectiles assuming that they are consistent. Expectiles are considered to be consistent if $m_{\alpha_t} > m_{\alpha_{t-1}}$. In order to have a well-conditioned problem we constrain the original problem and require smoothness. The conditions

$$\begin{aligned}\sum_i \varphi_i \psi(y_i, \alpha_t) &= \sum_i \varphi_i \alpha_{it} = 0 \qquad \forall t \\ \sum_i \varphi_i &= 1\end{aligned}$$

have to hold with φ a discrete approximation to the density f that we are looking for. The smoothness of φ is controlled by a third order penalty. We minimize the penalized least squares problem

$$S_f = \sum_{t=1}^{T+1} (h_t - \sum_i \alpha_{it}\varphi_i)^2 + \lambda_f \sum_i (\Delta^3 \varphi_i)^2$$

with $\boldsymbol{h}$ a vector of length $T+1$ and $\boldsymbol{h} = (0, \ldots, 0, 1)$ (details on the computation can be found in Schnabel and Eilers (2013a)).

This is the basic concept for estimating a density from a set of expectiles independent of the estimation method. At every point of interest of the independent variable, a conditional density can be determined. Estimating the density from a bundle model is especially straightforward as only one density based on c_α has to be estimated. This estimate is then shifted and scaled over the independent variable due to the used location-scale-model.

The bundle density can be estimated using the function `cdf.bundle` which uses as input an object of class `bundle` estimated with `expectreg.ls(..., estimate='bundle')`. The cdf, density and quantiles are calculated as above based on the vector $c_{\alpha_1}, \ldots, c_{\alpha_T}$ and returned.

```
R> exp.b <- expectreg.ls(hgt ~ rb(age, "pspline"),
      data=dutchboys[1:200,],estimate="bundle",
      expectiles=c(0.00001, seq(0.01,0.99, by=0.01),0.99999))
R> cdf.b <- cdf.bundle(exp.b)
```

Quantile bundle

The results from the bundle model and the estimated density enable us to infer a smooth non-crossing set of quantiles. We choose a set of quantiles with the help of the density and express it on the scale of c_α. By using the location-scale model and the corresponding values of c_α we can calculate quantile curves. The resulting quantiles are equally smooth and do not show crossovers similar to expectiles from the same bundle model. This approach is implemented in the function `quant.bundle`.

6.5 Semiparametric expectile regression

In an additive regression setting, we attempt to explain a response variable through the sum of a number of covariate effects. This leads to the regression model for d covariates

$$\boldsymbol{y} = \beta_0 + \beta_1 \boldsymbol{x}_1 + \ldots + \beta_{\tilde{d}} \boldsymbol{x}_{\tilde{d}} + f_1(\boldsymbol{x}_{\tilde{d}+1}) + \ldots + f_{d-\tilde{d}}(\boldsymbol{x}_d) + \boldsymbol{\epsilon}_\alpha,$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\tilde{d}}$ is the parametric part and $\boldsymbol{x}_{\tilde{d}+1}, \ldots, \boldsymbol{x}_d$ the nonparametric part. In order to implement this model in our expectile regression functions, we have to cast all effects into a unifying framework. The vector of function evaluations $\boldsymbol{f}_j$ is represented as the product of a design matrix $\boldsymbol{B}_j$ and a vector of regression coefficients $\boldsymbol{u}_j$ such that $\boldsymbol{f}_j = \boldsymbol{B}_j \boldsymbol{u}_j$. The complete predictor η can then be written as

$$\boldsymbol{\eta} = \boldsymbol{\beta X} + \boldsymbol{B}_1 \boldsymbol{u}_1 + \ldots + \boldsymbol{B}_{d-\tilde{d}} \boldsymbol{u}_{d-\tilde{d}}$$

with $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{\tilde{d}})'$ and $\boldsymbol{X} = (\mathbb{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\tilde{d}})$. Associated with each vector of regression coefficients is a quadratic penalty $\lambda_j \boldsymbol{u}_j' \boldsymbol{P}_j \boldsymbol{u}_j$ that enforces smoothness of the function $\boldsymbol{f}_j$.

In order to use `lsfit` we combine the effects in the following way: Matrix $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{d-\tilde{d}})$ and vector $\boldsymbol{v} = (\boldsymbol{\beta}', \boldsymbol{u}_1', \ldots, \boldsymbol{u}_{d-\tilde{d}}')'$ are the design matrix and regression coefficients corresponding to the model and $\boldsymbol{P} = \mathrm{diag}(\boldsymbol{0}, \lambda_1 \boldsymbol{P}_1, \ldots, \lambda_{d-\tilde{d}} \boldsymbol{P}_{d-\tilde{d}})$ is the complete penalty matrix. The reparameterization by Fahrmeir, Kneib, and Lang (2004) is applied as a default to center the bases around zero since the combination of multiple design matrices which contain intercepts would result in a collinear matrix $\boldsymbol{Z}$ and therefore in an ambiguous regression. The reparameterization has the great advantage that it automatically ensures identifiability of the functions $f_1(.), \ldots, f_{d-\tilde{d}}(.)$. The function `expectreg.qp` handles multiple covariates in the same way as `expectreg.ls`.

Besides penalized splines and parametric effects, there are a number of different effects that can be cast into this framework. The function `rb`, that constructs regression bases to fit the remaining package functions, allows for a wide variety of covariate modeling like ridge or random effects. Additionally, construction algorithms for spatial effects like tensor-product splines, Markov random fields and kriging are included and will be presented in the following.

Covariate effects	Option	Type	Boosting analogue
splines	`pspline`	nonlinear	`bbs`
parametric	`para`	categorical	`bols`
bivariate splines	`2dspline`	spatial	`bspatial`
Markov random field	`markov`	spatial	`bmrf`
radial basis	`radial`	spatial	-
kriging	`krig`	spatial	`brad`
ridge	`ridge`	categorical	`bridge`
random	`random`	random	`brandom`
user defined	`special`	-	`buser`

Table 6.2: Possible covariate effects in an expectile regression.

6.5.1 Spatial effects

Bivariate P-splines

The idea of P-splines can be extended to smoothing bivariate surfaces $f(x_1, x_2)$ by considering the tensor product basis consisting of all pairwise products $B_{k_1k_2}(x_1, x_2) = B_{k_1}(x_1) \cdot B_{k_2}(x_2)$ of univariate B-spline bases $B_{k_1}(x_1)$ and $B_{k_2}(x_2)$ in x_1- and x_2-direction, respectively (with $k_1 \in \{1, \dots, K_1\}$ and $k_2 \in \{1, \dots, K_2\}$). This yields the representation of a bivariate surface as

$$f(x_1, x_2) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} u_{k_1k_2} B_{k_1k_2}(x_1, x_2).$$

Therefore the design matrix consists of all evaluations of bivariate tensor product basis functions. The vector of regression coefficients is a vectorized version of the coefficient field $u_{k_1k_2}$. The penalty matrix is constructed from the Kronecker sum of the univariate penalty matrices, i. e. $\boldsymbol{P} = \boldsymbol{I}_{K_2} \otimes \boldsymbol{P}_1 + \boldsymbol{P}_2 \otimes \boldsymbol{I}_{K_1}$ where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are penalties for the univariate bases that have been employed for forming the tensor product.

Markov random fields

For discrete spatial data based on regional information $s \in \{1, \dots, S\}$, the vector of regression coefficients collects all distinct spatial effects, i.e., $\boldsymbol{\beta}_j = (\beta_{j,1}, \dots, \beta_{j,S})'$ where $\beta_s = f(s)$ is the spatial effect in region s. The design matrix then simply connects an observation i with the corresponding spatial effect $f(s_a)$, yielding

$$\boldsymbol{Z}[i,a] = \begin{cases} 1 & \text{if } y_i \text{ was observed in region } s_a, \\ 0 & \text{otherwise.} \end{cases}$$

The penalty matrix is chosen such that spatially adjacent regions share similar effects. This can be achieved by defining the penalty matrix $\boldsymbol{P}$ as an adjacency matrix

$$\boldsymbol{P}[a,b] = \begin{cases} -1 & a \neq b, s_a \sim s_b, \\ 0 & a \neq b, s_a \nsim s_b, \\ \omega_{s_u} & a = b, \end{cases} \tag{6.10}$$

where $s \sim r$ denotes that the two regions s and r are neighbors and ω_s denotes the total number of neighbors for region s. In the canonical case, two regions will be called neighbors if they share the same border coordinates. But also the user is able to define the neighborhood structure by passing the matrix $\boldsymbol{P}_j$ to the function. In a stochastic interpretation, the resulting penalty is equivalent to the assumption that $\boldsymbol{\beta}$ follows a Gaussian Markov random field (see Rue and Held, 2005 for details).

The neighborhood structure can either be given to the function `rb` directly in the form of the penalty matrix or by a complete map in the boundary format `bnd`. This is an `S3` class introduced by the package `BayesX` (Kneib, Heinzl, Brezger, and Sabanés Bové, 2013). An object from the class `bnd` contains a list of polygons represented as a matrix of coordinates. `expectreg` uses the method `bnd2gra` from the `BayesX` package to convert the polygon list into an adjacency matrix as defined in Equation (6.10). As output method the function `drawmap` is also borrowed from the same package. It provides the possibility to colour each polygon according to a given value. Therefore a vector with the fitted regression values of that spatial covariate is handed to the function, matching the order of the polygons / regions.

Radial bases / kriging

Continuous spatial effects can be approximated by tensor product spline bases. However, there might be spatial effects that, similar to the regions, do not fit with the rectangle that is the support of a bivariate spline basis. In that case it might be sensible to switch to a radial basis or a spatial basis based on kriging. The knots for these bases are chosen as a subset $\boldsymbol{k}_1, \dots, \boldsymbol{k}_K$ from the covariate observations $\boldsymbol{z}_{j1}, \dots, \boldsymbol{z}_{jn}$. The value of the basis $\boldsymbol{Z}$ for an arbitrary point $\boldsymbol{x}$ is then computed as a function $B_{\boldsymbol{k}}(r)$ of the Euclidean distance $r = ||\boldsymbol{k} - \boldsymbol{x}||$. A common choice for radial basis functions are thin plate splines that are defined as $B_{\boldsymbol{k}}(r) = r^2 \ln(r)$. The penalty matrix $\boldsymbol{P} = (B_{\boldsymbol{k}_i}(||\boldsymbol{k}_i - \boldsymbol{k}_j||))_{i,j}$ then comprises the evaluated distances between the knots. The same holds for the penalty matrix of kriging effects. Only the basis functions are derived from correlation functions. A popular model is defined by Matérn correlation functions like $B_{\boldsymbol{k}}(r, \phi) = \exp(-|r/\phi|)(1 + |r/\phi|)$ for a fixed $\phi > 0$. All these options are available in the R function `rb` for construction of a basis corresponding to the introduced effects. The value of ϕ is chosen proportional to $\max_{i,j}(||\boldsymbol{k}_i - \boldsymbol{k}_j||)$. For example,

```
R> b <- rb(x,type="krig",center=FALSE)
```

will construct an evaluated kriging basis from a two-dimensional covariate x according to the definitions above. The basis knots are chosen randomly from the observations with a number of $\min(50, n)$. The basis object `b` contains the evaluated basis matrix $\boldsymbol{Z}$, the penalty matrix $\boldsymbol{P}$ and the value for ϕ. In addition the values of `type` and `center` are returned.

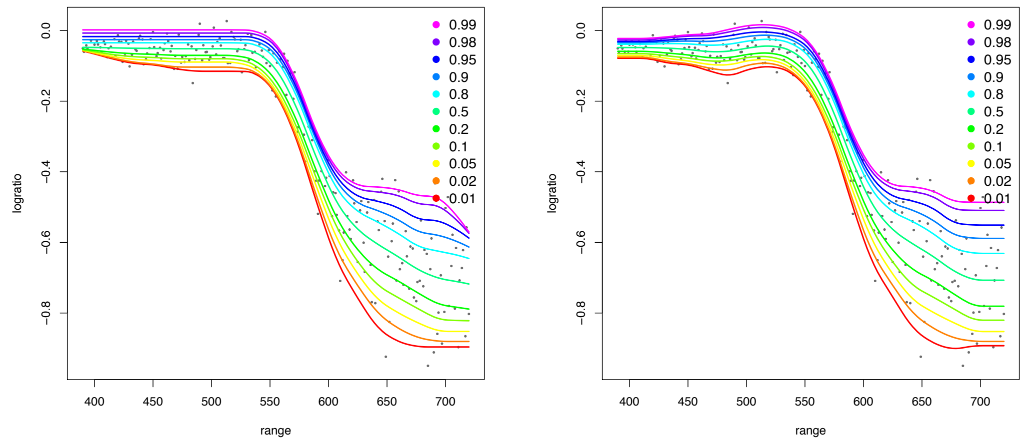
6.5.2 Shape-constrained P-splines

The function `mono` presents an alternative to the penalized spline bases defined by `rb`. Though constructed in the same way as described in Section 6.3.1, an additional penalty matrix $\boldsymbol{P}_c$ is added. The used penalty depends on the desired functional form to be estimated. The available possibilities for the use of the function `mono` are depicted in Table 6.3. Note that the reparameterization described before cannot be applied to this basis since the locality and neighbourhood of the splines has to be used.

Constraint	Penalty
increasing	positive first differences
decreasing	negative first differences
convex	positive second differences
concave	negative second differences
flatend	local first differences

Table 6.3: Available constraints $\boldsymbol{P}_c$ that can be added to a penalized spline basis.

In the least squares estimate from `expectreg.ls` an additional iteration is added where the constraint penalty $\lambda_c \mathbb{1}(\boldsymbol{P}_c \hat{\boldsymbol{u}}_\alpha^{[b-1]} < 0)\boldsymbol{P}_c$ is computed and added to the estimate ((6.5)) in the next iteration step. This is repeated until all constraints are met. The method `expectreg.qp` already includes a constraint matrix $\boldsymbol{A}$ for monotonicity in α. The constraint matrix $\boldsymbol{P}_c$ is simply appended to $\boldsymbol{A}$ for each expectile. The package `mboost` also supplies a baselearner with monotonicity constraints named `bmono`. In the boosting iterations the selected effect with the steepest gradient is only then included in the model if it meets the constraints given by `bmono`.



(a) `mono(range,"decreasing")` (b) `mono(range,"flatend")`

Figure 6.4: Expectile curves estimated for the lidar data using `expectreg.ls`, monotonously decreasing splines or flat effects at both ends of the covariates' support.

```
R> ex = expectreg.ls(logratio ~ mono(range,"decreasing"),
                                 data=lidar)
```

6.6 Examples

6.6.1 Childhood malnutrition

In the following example we analyse data on childhood malnutrition in India by using expectiles. The data set contains 4000 observations in six variables. For each child we have the subregion from India that it lives in, the age and body mass index of the child as well as the mother and a malnutrition score for the child called "stunting". The latter is of major interest here and has been analysed using quantile boosting by Fenske, Kneib, and Hothorn (2011). First, we perform a univariate analysis of this variable by computing a set of expectiles and displaying an E-E Plot to compare the distribution of "stunting" to a normal distribution.

Then we compute a geoadditive expectile regression with stunting as response, the BMI and the age of the child as nonlinear covariates and the region of India that the child lives in as spatial effect. The latter is modeled as a Markov random field and the spatial structure is passed to the function via the boundary data "india.bnd". Since the calculation of the cross-validation score is independent from the number of covariates, the smoothing parameters are optimized simultaneously. Finally we chose a specific set of four expectiles to be calculated. Due to the large data set and the high number of regions in India, these calculations will take a few minutes. Both the data set and the regional data from India are included in the package.

```
R> data("india")
R> data("india.bnd")
R> eenorm(india$stunting)
R> expectile(india$stunting)
        0      0.25       0.5      0.75         1
-599.0000 -247.9259 -175.4113 -102.9477  564.0000

R> ex <- expectreg.ls(stunting ~ rb(cbmi) + rb(cage) +
                       rb(distH,type="markov",bnd=india.bnd),
                       data=india,estimate="laws",smooth="acv",
                       expectiles=c(0.05,0.2,0.8,0.95))
R> plot(ex)
```

We can see from Figure 6.5 that it is not unlikely that the marginal distribution of stunting follows a normal distribution. The function `expectile` returns some key expectiles that also show a symmetry between the 0.25 and the 0.75 expectile. The results of the additive regression are displayed in Figures 6.6 and 6.7. The conditional distribution of stunting is plotted separately for each covariate. The curves show both a general trend along the covariate and the changes in variation of the response. Hence we can see that malnutrition gets worse up to an age of 24 months and that children with a very low BMI of 10 or with a BMI above 20 are affected worst. As a final analysis we can put together a new data set with 10 observations that combine both risks and predict the stunting values for those observations and the different expectiles.

6.6.2 Dutch growth data

As a second example, we consider the Dutch Growth Data (van Buuren and Fredriks, 2001). We apply two different estimators for expectiles to the data. With the resulting objects we can now carry out the two available distribution estimations introduced in Section 6.4. Therefore the regressions are computed for an especially dense set of expectiles. For the

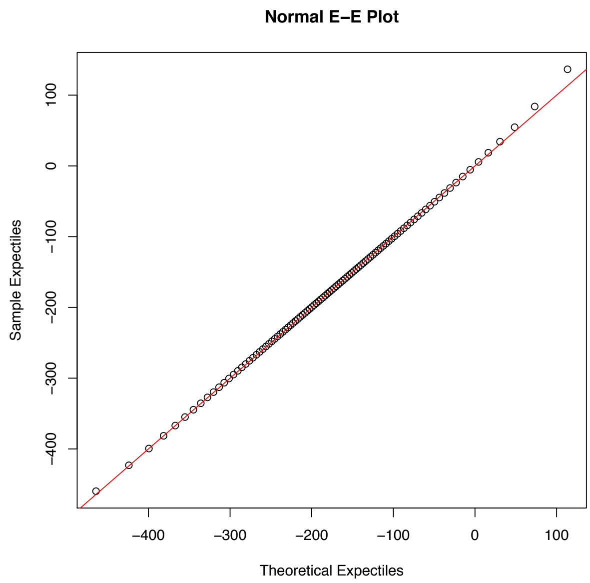
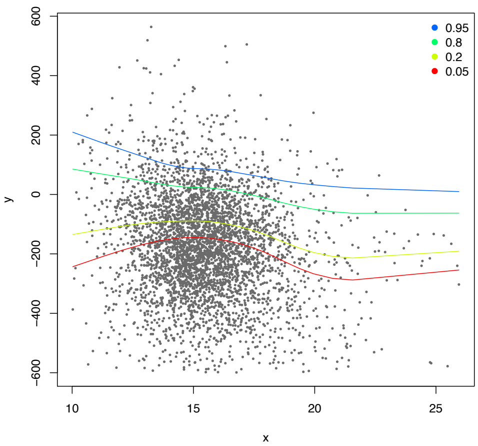
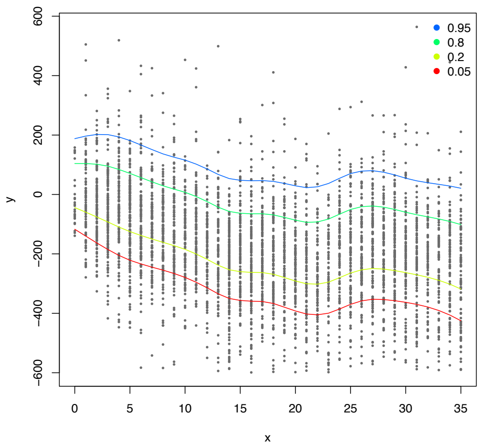


Figure 6.5: E-E Plot for the response variable “stunting” from the childhood malnutrition data.

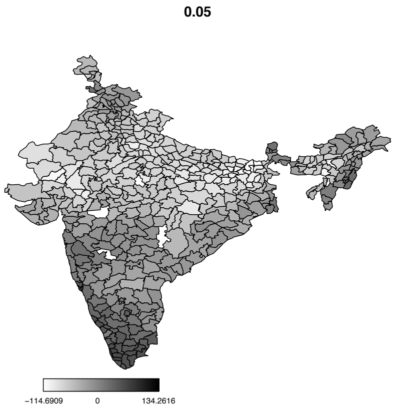


(a) `rb(cmbi)` - BMI

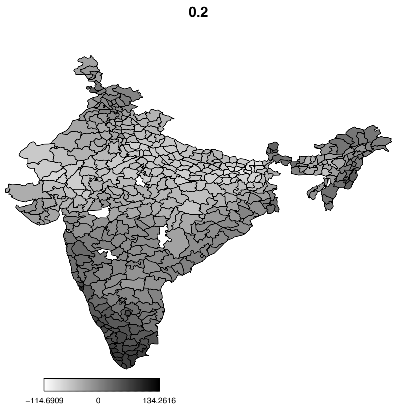


(b) `rb(cage)` - Age

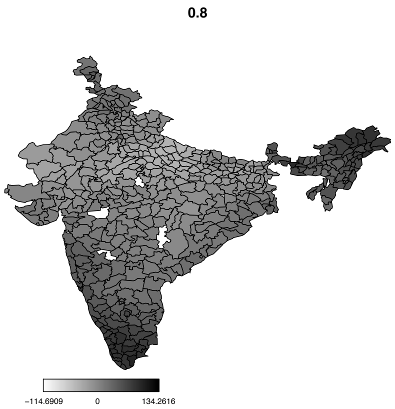
Figure 6.6: Results from the geoadditive expectile regression for the data on childhood malnutrition in India. Effects of the nonlinear variables BMI of child and age of child in months on the response variable stunting are shown, each for four expectiles.



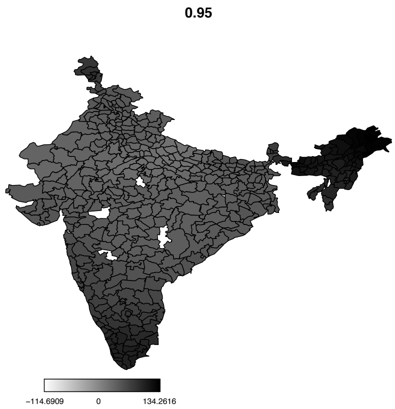
(a) $\alpha = 0.05$



(b) $\alpha = 0.2$



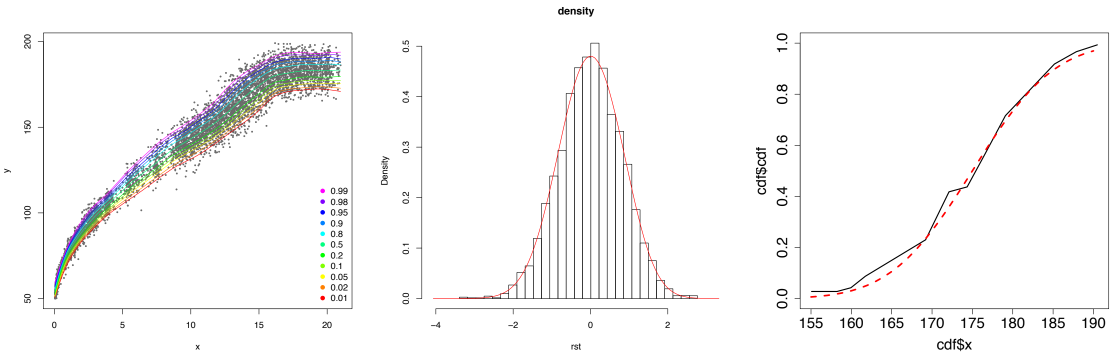
(c) $\alpha = 0.8$



(d) $\alpha = 0.95$

Figure 6.7: Results from the geoadditive expectile regression for the data on childhood malnutrition in India. Effects of the spatial effect of the regions of India on the response variable stunting are shown, each for four expectiles. Obtained with `rb(distH,"markov",bnd=india.bnd)`.

expectile bundle we have the option `density=TRUE` such that expectiles are computed in steps of 0.01. The function `expectreg.qp` is provided with a custom vector of asymmetries. In Figure 6.8 we can then see the density based on the bundle asymmetries and the cumulative distribution function estimated at age 15. Both methods indicate a symmetric distribution with resemblance to a normal distribution that is also displayed as a comparison to the cdf.



(a) `expectreg.ls` (b) `cdf.bundle(ex2)` (c) `cdf.qp(ex3,x=15)`

Figure 6.8: Estimated expectiles (a), conditional density (b) and cumulative distribution function (c) based on expectile bundle and sheet estimate, respectively. The latter is conditioned on a covariate value of 15 and compared to a normal distribution (drawn in red).

```
R> data("dutchboys")
R> ex2 <- expectreg.ls(hgt ~ rb(age), data=dutchboys,
      estimate="bundle",smooth="schall",density=TRUE)
R> plot(ex2)
R> bun <- cdf.bundle(ex2)
R> exp <- c(0.001,0.005,0.01,seq(0.02,0.98,by=0.16),0.99,0.995,0.999)
R> ex3 <- expectreg.qp(hgt ~ rb(age), data=dutchboys, expectiles=exp)
R> cdf <- cdf.qp(ex3,x=15)
R> plot(cdf$x,cdf$cdf,type="l",ylim=c(0,1))
R> lines(155:190,pnorm(155:190,m=175,sd=8),col="red")
```

6.7 Conclusions / future work

With `expectreg` we have introduced the first comprehensive package for the use of expectiles. The package allows for the analysis of univariate distributions and data as well as highly flexible regression scenarios. A vector of observations can easily be tackled by estimating a set of expectiles in order to obtain a nonparametric estimate of the distribution. In most cases, and especially for small samples, estimating expectiles is more efficient than the use of quantiles. This has also been shown by Schulze Waltrup, Sobotka, Kneib, and Kauermann (2014). E-E plots also produce smoother results than their quantile counterparts while conveying similar information. These attributes make expectiles a compelling candidate in a regression scenario. They combine the flexibility in modeling and the efficiency of a least squares mean regression with the gain of information from quantile regression. The package `expectreg` provides interested R users with simple methods that have a similar structure as existing functions for univariate data analysis and semiparametric regression.

The package also provides a number of possible solutions for problems faced in expectile regression. A single expectile curve can be interpreted, but its interpretation is not as intuitive as the interpretation of a single quantile curve. This can be handled in different ways. First of all, by default a set of eleven expectiles is estimated in each regression. Hence, a relatively complete picture of the conditional distribution of the response is drawn and therefore a clear gain of information is visible. We also included the implementation of two approaches to compute quantiles from a set of expectiles which allows for the calculation of interpretable quantiles with an efficient estimate. As there might be issues with performing a quantile regression, we provide the possibility to obtain expectiles and subsequently calculate the interpretable quantiles from one of the two distribution estimation methods. This can be very helpful, since there are a number of disadvantages when regression quantiles are to be estimated in R. Since quadratic penalties are not easily combinable with the linear programming techniques used in quantile regression, the spatial effects described in this paper are not available.

Another issue that mainly affects quantile regression is the problem of crossing estimates. While the estimates should be monotonously increasing with the chosen asymmetry, it is possible in practice that quantiles as well as expectiles cross due to sampling variation. However, simulations have shown that expectiles are less likely to cross than quantiles (see Schulze Waltrup, Sobotka, Kneib, and Kauermann, 2014). In addition, for quantiles and expectiles non-crossing estimates have been proposed and a non-crossing expectile estimation procedure is included in `expectreg`. An implementation of non-crossing quantiles can be found in the R package `quantregGrowth` (see Muggeo, Sciandra, Tomasello, and Calvo, 2013).

Despite all the functionality that is already available, we do not regard `expectreg` as finished. As mentioned before, the restricted expectiles and the bundle are yet to be implemented with the boosting algorithm. In the more distant future we aim on including interactions in the models in order to model spatio-temporal effects. A follow-up package including risk measures based on expectiles, like the expected shortfall (ES) is also planned.

Acknowledgements

The authors would like to thank Paul Eilers for his insight and help with both the software and this paper as well as Torsten Hothorn for his valuable comments during the development of the package. Financial support from the German Research Foundation (DFG) grant KN 922/4-1 is gratefully acknowledged.

Chapter 7

Conclusion

7.1 Summary

This thesis started with an introduction to semiparametric expectile modeling. We also offered an interpretation for expectiles in Section 2.4. Additionally there is the possibility to calculate quantiles from expectiles. In Chapter 3 we used the one-to-one connection between expectiles and the distribution function and quantiles and the distribution function to calculate quantiles from a set of expectiles. Simulations indicate, that there is no loss in efficiency. We saw, that the connection between quantiles and expectiles and the computation of quantiles from expectiles became extremely useful when using the expected shortfall as risk measure. Although estimated regression expectiles tend to cross less often than regression quantiles, crossing of neighboring expectile curves may occur. Therefore we introduced a method to estimate non-crossing expectiles in Chapter 4. We also saw that the L_2 nature of expectiles easily allows for common generalizations known from mean regression as the incorporation of a random intercept within an additive framework. The method was accompanied by an application: The random intercept was motivated by panel observations from the German Socio Economic Panel data. In Chapter 5 the estimation of expectiles within the context of unequal probability sampling was introduced. We used the expectiles to estimate the distribution function and its inversion which led to quantiles. The quantiles from expectiles were compared with benchmark methods. An implementation of the methods can be found within the open source software (R Core Team, 2014). The R package `expectreg` by Sobotka, Schnabel, and Schulze Waltrup (2013) is a collection of the implementation of numerous methods regarding expectiles. The function for the estimation of quantiles from a set of expectiles and the additive expectile regression with a random intercept are part of the R package `expectreg` (next to many other methods). A description of the R package `expectreg` was given in Chapter 6.

All in all, we saw during this thesis that due to its L_2 nature expectile regression is easily extended and generalized in the same way as mean regression. The L_2 nature is also the reason why the result of an expectile regression, as well as for mean regression, is influenced by outliers.

7.2 Discussion and Outlook

As each chapter is provided with an own discussion we focus on some points which have not been mentioned yet.

Although there exist first extensions and generalizations of expectile regression further research needs to be done. Model selection and variable selection for example have not been considered yet. A first approach to variable selection is provided by the asymptotic confidence intervals described in Sobotka, Kauermann, Schulze Waltrup, and Kneib (2013), but to our knowledge measures for the goodness of fit of an expectile model have not been evaluated yet. The estimation of quantiles from expectiles described in Chapter 3 relies on the assumption of a piecewise linear density and therefore may be further optimized. Also the mixed expectile model described in Chapter 4 needs to be extended further. The non-crossing condition presented in the same chapter is a rather loose one and other, more strict non-crossing conditions are easily integrated as long as they can be formulated as a linear constraint on the coefficients. In fact, as we are facing a quadratic optimization problem in expectile regression, all kinds of constraints regarding e.g. shape or monotonicity can be incorporated rather quick as long as they can be formulated as linear constraint on the coefficients. A numerical issue concerning expectiles is the computation time. When estimating a set of many expectiles for a large dataset the computation of expectile estimates, especially for expectile sheets as described in Chapter 4, can be time consuming. One possibility to speed things up could be to use an array formulation as proposed in Currie, Durbán, and Eilers (2006) (see also Schnabel and Eilers, 2014).

Maybe the most important task for further research is to even more encourage and distribute the use of expectiles. The argument most often used against expectiles is their lack of interpretability. As we saw in the introductory Chapter 2 and also in the remainder of this thesis, there are many approaches offering an interpretation of expectiles and also the interpretation of a single expectile curve is possible. We hope that this thesis will make a small contribution to the spread of expectile regression as data analysis tool.

Appendix A

Notation

α	asymmetry parameter
m_α	α–expectile
$m(\alpha, x)$	expectile sheet
q_α	α–quantile
$q(\alpha, x)$	quantile sheet
n	sample size
y	response
$\boldsymbol{y}$	response vector
x	covariate
$\tilde{d}$	number of linear effects
$\tilde{d}+1, \ldots, d$	index of nonlinear effects
$\boldsymbol{x}$	covariate vector
$\boldsymbol{X}$	parametric design matrix
$\boldsymbol{\beta}$	vector of parametric coefficients
$\boldsymbol{B}$	B-spline basis matrix
$\boldsymbol{u}$	vector of coefficients for B-spline basis
$\boldsymbol{Z}$	general basis matrix
$\boldsymbol{v}$	general vector of coefficients
$\boldsymbol{U}$	design matrix of random effects
$\boldsymbol{\gamma}$	vector of coefficients for random effects
ϵ	residuals

σ^2_ϵ	variance of errors
σ^2_γ	variance of random intercept
ρ_α	check function
w_α	weights
$\boldsymbol{W}$	matrix of weights
$\boldsymbol{D}$	matrix of differences
$\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{D}$	penalty matrix
λ	penalty parameter
$\boldsymbol{P} = \lambda\boldsymbol{K}$	penalty matrix
η	linear predictor
$\boldsymbol{A}$	constraint matrix enforcing non-crossing conditions
$F(.)$	cumulative distribution function
$f(.)$	probability density function
$G(.)$	partial moment function
$h(.)$	transfer function between expectiles and quantiles
$\hat{\zeta}$	estimated steps of distribution function
π	probability of inclusion

Appendix B

Appendix to Chapter 3

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

We assume that y has finite second moments, then with (3.3)

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

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

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

Hence $\tilde{\alpha} < \alpha$ for $\alpha \to 0$. Note that since $f(q_\alpha) = o\big(|q_\alpha|^{-3}\big)$ which follows due to the existence of second order moments, we find again that nominator and denominator of (B.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 (B.2), that $\lim\limits_{\alpha\to 0} \alpha/\tilde{\alpha} = \mathit{const} \geq 1$, while if $f(q_\alpha)$ is proportional to $|q_\alpha|^{-(3+\delta)}$ for some $\delta > 0$ we get with the same arguments that $\lim\limits_{\alpha\to 0} \alpha/\tilde{\alpha} = \infty$.

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 equation (3.4) 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

$$m_{0.5} = \int_{-\infty}^{\infty} y \; dF(y) = \frac{m_L + m_{L+1}}{2} + \sum_{l=1}^{L} (c_l - c_{L+1})\zeta_l \tag{B.3}$$

with the approximation of $F(.)$ from section 3.2.1. Remembering the definition of the expectiles (3.4), we define function $g_l(.)$ by

$$g_l(\zeta_l, \ldots, \zeta_1) = m_l - \frac{(1-\alpha_l)G_l(\zeta_l, \ldots, \zeta_1) + \alpha_l(m_{0.5} - G_l(\zeta_l, \ldots, \zeta_1))}{(1-\alpha_l)F_l(\zeta_l, \ldots, \zeta_1) + \alpha_l(1 - F_l(\zeta_l, \ldots, \zeta_1))} \quad \text{for } l = 1, \ldots, L. \tag{B.4}$$

We now need $\zeta_1, \ldots, \zeta_L$ 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 $g_l(\zeta)$ under certain restrictions: We face the minimization problem

$$\min_{\zeta_1,\ldots,\zeta_L} S(\zeta_1, \ldots, \zeta_L) = \min_{\boldsymbol{\zeta_1,\ldots,\zeta_L}} \sum_{l=1}^{L} (g_l(\zeta_l, \ldots, \zeta_1)^2) \tag{B.5}$$

under the constraints that $\zeta_l \geq 0$ and $\sum_{l=1}^{L} \zeta_l \leq 1$ which is solved by Newton's method in optimization and also implemented in the R package "expectreg" by Sobotka, Schnabel, and Schulze Waltrup (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 5 times the squared empirical variance of the data (for each of the three distributions considered and for both sample sizes $n = 199$ and $n = 499$).

Appendix C

On Confidence Intervals for Semiparametric Expectile Regression

---

This manuscript is developed by Fabian Sobotka, Göran Kauermann and Thomas Kneib and in minor parts by Linda Schulze Waltrup. It is appeared as *On confidence intervals for semiparametric expectile regression* in Statistics and Computing. The asymptotic results were derived in joint work by Göran Kauerman, Thomas Kneib, Fabian Sobotka and Linda Schulze Waltrup. The corresponding sections were written by Göran Kauermann and Thomas Kneib. Fabian Sobotka implemented the confidence intervals, ran the simulation study and applied the results within the example. All authors contributed to the general investigation of the scientific problem and were involved in writing and proofreading the manuscript.

Abstract

In regression scenarios there is a growing demand for information on the conditional distribution of the response beyond the mean. In this scenario quantile regression is an established method of tail analysis. It is well understood in terms of asymptotic properties and estimation quality. Another way to look at the tail of a distribution is via expectiles. They provide a valuable alternative since they come with a combination of preferable attributes. The easy weighted least squares estimation of expectiles and the quadratic penalties often used in flexible regression models are natural partners. Also, in a similar way as quantiles can be seen as a generalization of median regression, expectiles offer a generalization of mean regression. In addition to regression estimates, confidence intervals are essential for interpretational purposes and to assess the variability of the estimate, but there is a lack of knowledge regarding the asymptotic properties of a semiparametric expectile regression estimate. Therefore confidence intervals for expectiles based on an asymptotic normal distribution are introduced. Their properties are investigated by a simulation study and compared to a bootstrap-based gold standard method. Finally the introduced confidence intervals help to evaluate a geoadditive expectile regression model on childhood malnutrition data from India.

C.1 Introduction

C.1.1 Expectiles

Recent interest in modern regression modeling has focused on extending available model specifications beyond mean regression by describing more general properties of the response distribution. For example, Rigby and Stasinopoulos (2005) proposed regression models for location, scale and skewness where separate predictors can be specified for various parameters of a response distribution. A completely distribution free approach is quantile regression (Koenker and Bassett, 1978) where regression effects on the conditional quantile function of the response are assumed. Combining models for a large set of quantiles then allows to characterize the complete conditional distribution of the response.

Quantile regression for the α-quantile with $\alpha \in (0, 1)$ relies on the regression specification

$$y_i = \eta_{i,\alpha} + \epsilon_{i,\alpha}, \quad i = 1 \ldots, n, \tag{C.1}$$

where $\eta_{i,\alpha}$ is a (quantile-specific) predictor and $\epsilon_{i,\alpha}$ are independent error terms. Instead of imposing the usual mean regression model assumption that $E(\epsilon_{i,\alpha}) = 0$, quantile regression relies on the assumption that for the quantile function Q holds that $Q_{\epsilon_{i,\alpha}}(\alpha) = 0$, i.e. the α-quantile of the error distribution is zero. This implies that the conditional quantile of the response y_i is given by the predictor $\eta_{i,\alpha}$. Note that no specific distribution is assumed for the error terms or responses and that in particular the error distribution may differ between individuals. Estimation of quantile specific predictors now relies on minimizing the asymmetrically weighted absolute residuals criterion $\sum_{i=1}^{n} w_{i,\alpha}|y_i - \eta_{i,\alpha}|$ with weights

$$w_{i,\alpha} = w_{i,\alpha}(\eta_{i,\alpha}, y_i) = \begin{cases} \alpha, & \text{for } y_i \geq \eta_{i,\alpha} \\ 1 - \alpha, & \text{for } y_i < \eta_{i,\alpha}. \end{cases} \tag{C.2}$$

A computationally attractive alternative to quantile regression is expectile regression, where absolute residuals are replaced with squared residuals yielding the fit criterion

$$\sum_{i=1}^{n} w_{i,\alpha}(y_i - \eta_{i,\alpha})^2$$

with weights as defined in (C.2). The underlying assumption in regression model (C.1) is that the α-expectiles m_α of the error terms are zero. They are implicitly defined by $m_\alpha =$

$\arg\min_m E\left[w_\alpha(m, y)(y-m)^2\right]$. Least asymmetrically weighted squares (LAWS) estimation of expectiles dates already back to Newey and Powell (1987) but recently re-gained interest in the context of semiparametric or geoadditive regression (see for example Schnabel and Eilers, 2009b; Sobotka and Kneib, 2012). Expectile estimation is thereby a special form of M-quantile estimation, see Breckling and Chambers (1988), Jones (1994). One of the advantages of expectile regression is that estimation basically reduces to (iteratively) weighted least squares fits since the optimality criterion is differentiable with respect to the regression effects while linear programming routines have to be used in case of quantile regression. This is of particular relevance when considering more flexible regression specifications as for example in geoadditive regression. The effects included here depend on a quadratic penalty for smooth estimates which can easily be included in a least squares estimation procedure. Further, when using expectiles (or quantiles) we try to get a complete picture of the conditional distribution of the response while at the same time avoiding a parametric specification for the distribution. To achieve this, we need to consider a set of expectiles or quantiles. In this scenario, a single estimate would not hold more information than the mean. Therefore we regard the reduced interpretability of expectiles as non-critical. Nevertheless, the estimation efficiency of expectiles and the interpretability of quantiles could be combined, if wished for, since Efron (1991) already proposed a method to obtain quantiles from a set of expectiles.

In summary, point estimates for expectile regression are easily derived for simple as well as complex models but their statistical properties are not yet well understood. In contrast, confidence intervals and significance tests for quantile regression have been studied extensively in the literature, relying for example on asymptotic considerations, the connection of quantiles to ranks or on bootstrap procedures (Koenker, 2005; Kocherginsky, He, and Mu, 2005; Buchinsky, 1998). In this paper, we derive asymptotic properties of expectile regression estimates and use them to construct corresponding confidence intervals. We continue the work of Newey and Powell (1987) by introducing a correction for the asymptotic results and extending them to semiparametric regression models. Further we determine the empirical properties of the asymptotic results. Therefore we state bootstrap-based confidence intervals as a computationally demanding gold standard for comparison with confidence intervals relying on asymptotic normality. Pointwise bootstrap percentile intervals have already been considered in Sobotka and Kneib (2012). However, the empirical properties were not determined and the method proved to be impractical for larger data sets due to the highly increased computational costs.

C.1.2 Geoadditive Expectile Regression

The need for our methodological innovations has arisen during a large-scale application on childhood malnutrition in developing countries where the impact of a large set of covariates should be assessed with respect to their impact on the nutritional status of children. Exploring not only the conditional mean but also extreme parts of the conditional distribution is of particular interest in this application since it allows to determine specific determinants of severe malnutrition by modeling lower expectiles. A comparable application is considered in Fenske, Kneib, and Hothorn (2011) who use boosting to estimate regression quantiles in a high-dimensional additive quantile regression model, but spatial effects could not be included and confidence intervals are not provided. For the assessment of estimation uncertainty they apply cross-validation in combination with the stability selection procedure recently proposed by Meinshausen and Bühlmann (2010). In this paper we use an extended, geoadditive model specification as introduced by Kammann and Wand (2003). The model definition combines parametric and nonlinear effects as well as spatial effects from geostatistics like kriging and can therefore be seen as a highly general semiparametric mixed model. For our application the geoadditive specification yields

$$\begin{aligned}\eta_{i,\alpha} \;=\;& (\text{csex},\ldots,\text{car})_i'\boldsymbol{\beta}_\alpha + f_{1,\alpha}(\text{cage}_i) + f_{2,\alpha}(\text{cfeed}_i) \\ & + f_{3,\alpha}(\text{mbmi}_i) + f_{4,\alpha}(\text{mage}_i) + f_{5,\alpha}(\text{medu}_i) \\ & + f_{6,\alpha}(\text{medupart}_i) + f_{\text{spat},\alpha}(\text{district}_i)\end{aligned} \tag{C.3}$$

where $\boldsymbol{\beta}_\alpha$ corresponds to parametric effects of categorical covariates such as gender of the child (csex) or household-specific asset indicators (e.g. presence of a car), $f_{1,\alpha},\ldots,f_{6,\alpha}$ are nonlinear effects of the continuous covariates age of the child in months (cage), duration of breastfeeding in months (cfeed), body mass index of the mother at birth (mbmi), age of the mother at birth (mage) and education years of the mother and the mother's partner (medu, medupart) modeled via penalized splines and $f_{\text{spat},\alpha}$ is a spatial effect corresponding to a Gaussian Markov random field.

The rest of this paper is structured as follows: Section C.2 presents results on the asymptotic normality of expectile regression estimates in simple parametric models and for semiparametric extensions relying on penalized estimation. Required nonlinear and spatial effects are introduced alongside. Section C.3 uses these asymptotic results to derive confidence intervals and also proposes bootstrap-based alternatives. Simulations and results for the childhood malnutrition data are presented in Section C.4. The final Section C.5 summarizes the findings.

C.2 Asymptotics for Least Asymmetrically Weighted Squared Error Estimates

In the following, we assume that n metric observations $y_1, \dots, y_n$ are given. For the underlying unknown distribution we require the existence of second moments. Further, all inverted matrices are assumed to have full rank.

C.2.1 Parametric Models

We start our considerations with a simple, parametric model $\eta_{i,\alpha} = \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha$ and study the asymptotic behaviour of

$$\hat{\boldsymbol{\beta}}_\alpha = \arg\min_{\boldsymbol{\beta}_\alpha} \sum_{i=1}^n w_{i,\alpha}(\boldsymbol{\beta}_\alpha)(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha)^2.$$

where $w_{i,\alpha}(\boldsymbol{\beta}_\alpha) := w_{i,\alpha}(\eta_{i,\alpha}, y_i)$. Let $\boldsymbol{\beta}_\alpha^0$ be the true parameter vector implicitly defined through

$$0 = \sum_{i=1}^n \left\{ (1-\alpha) \int_{-\infty}^{\boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0} \left(y - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0\right) f(y|\boldsymbol{x}_i) \mathrm{d}y \right. \qquad \text{(C.4)}$$
$$\left. +\alpha \int_{\boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0}^{\infty} \left(y - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0\right) f(y|\boldsymbol{x}_i) \mathrm{d}y \right\}.$$

To avoid complexities arising from the dependence of the weights on the parameter vector, let for the moment $w_{i,\alpha}^0 = w_{i,\alpha}\left(\boldsymbol{\beta}_\alpha^0\right)$ be the "true" weights and define $\hat{\boldsymbol{\beta}}_\alpha^0$ as the minimizer of

$$\hat{\boldsymbol{\beta}}_\alpha^0 = \arg\min_{\boldsymbol{\beta}_\alpha} \sum_{i=1}^n w_{i,\alpha}^0 (y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha)^2 \qquad \text{(C.5)}$$

which can easily be derived explicitly as

$$\hat{\boldsymbol{\beta}}_\alpha^0 = \left(\sum_{i=1}^n w_{i,\alpha}^0 \boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \left(\sum_{i=1}^n w_{i,\alpha}^0 \boldsymbol{x}_i y_i\right). \qquad \text{(C.6)}$$

Since the weights are considered as fixed we end up with standard weighted regression and obtain the following result:

Lemma 1. *The least asymmetrically weighted squares estimate with fixed weights is asymp-*

totically normal, i.e.

$$\hat{\boldsymbol{\beta}}_\alpha^0 \overset{a}{\sim} N\left(\boldsymbol{\beta}_\alpha^0, \mathrm{Var}(\hat{\boldsymbol{\beta}}_\alpha^0)\right). \tag{C.7}$$

with covariance matrix

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}_\alpha^0) = &\left(\sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left\{\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{x}_i' \,\mathrm{Var}\left(\varpi_{i,\alpha}^0 (y_i - \boldsymbol{x}_i' \boldsymbol{\beta}_\alpha)\right)\right\} \\ &\left(\sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}\end{aligned} \tag{C.8}$$

with $\varpi_{i,\alpha}^0 = E(w_{i,\alpha}^0) = (1-\alpha)P(y_i < \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0) + \alpha\, P(y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0)$.

Proof. With fixed weights, it is easy to show that

$$E(w_{i,\alpha}^0 y_i) = (1-\alpha)\int_{-\infty}^{\boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0} y f(y|\boldsymbol{x}_i)\mathrm{d}y + \alpha \int_{\boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0}^{\infty} y f(y|\boldsymbol{x}_i)\mathrm{d}y,$$

which, combined with the implicit definition of the expectile (C.4), yields

$$E\left(\sum_{i=1}^n w_{i,\alpha}^0 \boldsymbol{x}_i y_i\right) = \sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i' \boldsymbol{\beta}_\alpha^0$$

Applying standard expansion techniques to the weights in the first component in (C.6) yields

$$\left(\sum_{i=1}^n w_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} = \left(\sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} + O_p\left(n^{-1}\right),$$

so that we can extract the asymptotically leading components in (C.6) through

$$\begin{aligned}\hat{\boldsymbol{\beta}}_\alpha^0 - \boldsymbol{\beta}_\alpha^0 = &\left(\textstyle\sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1} \left(\textstyle\sum_{i=1}^n \varpi_{i,\alpha}^0 \boldsymbol{x}_i \left(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0\right)\right) \\ &+ O_p(n^{-1}).\end{aligned}$$

This shows that $E(\hat{\boldsymbol{\beta}}_\alpha^0) = \boldsymbol{\beta}_\alpha^0 + O(n^{-1})$ and the variance of the weighted least squares estimate with fixed weights equals (C.8). With the variance being of order $O(n^{-1})$, we obtain $\hat{\boldsymbol{\beta}}_\alpha^0 - \boldsymbol{\beta}_\alpha^0 = O_p(n^{-1/2})$ which, together with (C.9), yields the asymptotic normality (C.7). □

The next step in our consideration is to replace weights $w_{i,\alpha}^0 = w_{i,\alpha}(\boldsymbol{\beta}_\alpha^0)$ in (C.5) by its estimate $\hat{w}_{i,\alpha}^0 = w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha^0)$, that is we allow the weights to depend on the parameter estimate.

Theorem 1. *The least asymmetrically weighted squares estimate with estimated weights is asymptotically normal, i.e.*

$$\hat{\boldsymbol{\beta}}_\alpha \overset{a}{\sim} N\left(\boldsymbol{\beta}^0_\alpha, \mathrm{Var}(\hat{\boldsymbol{\beta}}^0_\alpha)\right). \tag{C.9}$$

A proof is available under the assumptions stated in the beginning. It is provided in the appendix and follows a similar line of thought as in Newey and Powell (1987). The inner component (C.8) of the variance in (C.7) and (C.9), respectively, can easily be derived analytically, but the analytic form is hard to estimate. We therefore suggest to replace $\mathrm{Var}(\varpi^0_{i,\alpha}(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}^0_\alpha))$ by its empirical version

$$(w^0_{i,\alpha})^2 \left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}^0_\alpha\right)^2. \tag{C.10}$$

Apparently, replacing (C.10) with its fitted version by substituting $\boldsymbol{\beta}^0_\alpha$ with its estimate $\hat{\boldsymbol{\beta}}^0_\alpha$ will lead to down-biased estimates since fitted squared expectile residuals underestimate the variance, like in classical regression. We therefore need to adjust (C.10) when applying its fitted version. From mean regression we already know that without further assumptions for the distribution of the residuals we have

$$\mathrm{Var}\big\{\big(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}^0_\alpha\big)\big\} = \mathrm{Var}\big\{\big(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}^0_\alpha\big)\big\}(1 - h_{ii})$$

with h_{ii} being the ith diagonal element of the hat matrix $\boldsymbol{H}$, say. For expectile regression we obtain the generalized hat matrix $\boldsymbol{H}^\alpha = \left(h^\alpha_{ij}\right)_{ij}$ with

$$h^\alpha_{ij} = w^0_{i,\alpha}\boldsymbol{x}_i'\left(\sum_{k=1}^n w^0_{k,\alpha}\boldsymbol{x}_k\boldsymbol{x}_k'\right)^{-1}\boldsymbol{x}_j,$$

that coincides with the OLS hat matrix for $\alpha = 0.5$, i.e. $\boldsymbol{H} = \boldsymbol{H}^{0.5}$. Therefore we use (C.10) but estimate the variance with the adjusted fitted residuals

$$(\hat{w}_{i,\alpha})^2 \frac{\left(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}^0_\alpha\right)^2}{1 - h^\alpha_{ii}}, \tag{C.11}$$

where $\hat{w}_{i,\alpha} = w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha)$.

C.2.2 Semiparametric Models

Now we extend the results from the previous section to semiparametric regression models with generic predictor

$$\eta_{i,\alpha} = \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha + \sum_{j=1}^{d-\tilde{d}} f_{j,\alpha}(x_{i(\tilde{d}+j)}) = \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha + \sum_{j=1}^{d-\tilde{d}} \boldsymbol{b}_{ij}'\boldsymbol{u}_{j,\alpha}$$

where $\boldsymbol{x}_i'\boldsymbol{\beta}_\alpha$ summarizes usual parametric, linear effects of $x_1, \dots, x_{\tilde{d}}$ while $f_{1,\alpha}(.), \dots, f_{d-\tilde{d},\alpha}(.)$ represent generic semiparametric effects of covariates $x_{\tilde{d}+1}, \dots, x_d$. These may for example stand for nonlinear effects of continuous covariates or spatial effects as in our application (also compare equation (C.3)) but may also correspond to more complex terms such as varying coefficients or interaction surfaces (see Fahrmeir, Kneib, and Lang, 2004, for more details on available model terms). Each of the generic regression terms can then be expanded in terms of basis functions, yielding a representation as $f_{j,\alpha}(x_{i(\tilde{d}+j)}) = \boldsymbol{b}_{ij}'\boldsymbol{u}_{j,\alpha}$ where $\boldsymbol{b}_{ij}$ comprises the basis function evaluations while $\boldsymbol{u}_{j,\alpha}$ is a vector of basis coefficients.

To enforce specific properties of the resulting estimates such as smoothness, estimation then typically relies on penalized fit criteria. In case of expectile regression, this yields

$$\sum_{i=1}^{n} w_{i,\alpha}(\eta_{i,\alpha})(y_i - \eta_{i,\alpha})^2 + \sum_{j=1}^{d-\tilde{d}} \lambda_{j,\alpha}\boldsymbol{u}_{j,\alpha}'\boldsymbol{K}_j\boldsymbol{u}_{j,\alpha}.$$

where $\lambda_{j,\alpha} \geq 0$, $j = 1, \dots, d - \tilde{d}$ are smoothing parameters and $\boldsymbol{K}_j$ are appropriate penalty matrices.

The two relevant examples of semiparametric model terms in the context of our application are penalized splines and Gaussian Markov random fields. The former enables estimation of nonlinear effects $f_{j,\alpha}(x_{\tilde{d}+j})$ of a single continuous covariate $x_{\tilde{d}+j}$ and relies on a basis expansion in terms of B-splines in combination with a difference penalty for the basis coefficients. Therefore, in this case $\boldsymbol{b}_{ij}' = (B_1(z_i), \dots, B_K(z_i))$ where $B_1, \dots, B_K$ is a K-dimensional B-spline basis and $\boldsymbol{K}_j = \boldsymbol{D}'\boldsymbol{D}$ with a difference matrix $\boldsymbol{D}$. The penalty $\boldsymbol{u}_{j,\alpha}'\boldsymbol{K}_j\boldsymbol{u}_{j,\alpha}$ then consists of the sum of all squared differences of adjacent coefficient sequences and penalizes large variation in the function estimate (compare Eilers and Marx, 1996). Gaussian Markov random fields allow to estimate spatial effects based on geographical data. Suppose that each individual observation pertains to one region s_i from a fixed set of regions $\mathcal{S} = \{1, \dots, S\}$. Then the design vector $\boldsymbol{b}_{ij}$ is an S-dimensional indicator vector with a one at the position of the region of observation i and zeros otherwise while the vector of coefficients $\boldsymbol{u}_{j,\alpha}$ simply collects all potential spatial effects. The penalty matrix should enforce spatial smoothness and therefore has the structure of an adjacency matrix such that the penalty $\boldsymbol{u}'_{j,\alpha}\boldsymbol{K}_j\boldsymbol{u}_{j,\alpha}$ consists of all squared differences between spatial effects of neighboring regions (see Rue and Held, 2005, for details).

In any case, the estimates in semiparametric expectile regression models for fixed smoothing parameters can always be written as

$$\hat{\boldsymbol{v}}_\alpha = \left(\sum_{i=1}^n \boldsymbol{z}'_i w_{i,\alpha}\boldsymbol{z}_i + \boldsymbol{P}\right)^{-1}\left(\sum_{i=1}^n \boldsymbol{z}'_i w_{i,\alpha} y_i\right)$$

where $\boldsymbol{v}_\alpha = (\boldsymbol{\beta}'_\alpha, \boldsymbol{u}'_{1,\alpha}, \ldots, \boldsymbol{u}'_{d-\tilde{d},\alpha})'$ and $\boldsymbol{z}_i = (\boldsymbol{x}'_i, \boldsymbol{b}'_{i1}, \ldots, \boldsymbol{b}'_{i(d-\tilde{d})})'$ collect all regression coefficients and design vectors, respectively, and $\boldsymbol{P} = \operatorname{blockdiag}(\boldsymbol{0}, \lambda_{1,\alpha}\boldsymbol{K}_1, \ldots, \lambda_{d-\tilde{d},\alpha}\boldsymbol{K}_{d-\tilde{d}})$ is the complete penalty matrix.

Theorem 2. *For fixed smoothing parameters, the penalized least asymmetrically weighted squares estimate is asymptotically normal, i.e.*

$$\hat{\boldsymbol{v}}_\alpha \overset{a}{\sim} N\left(\boldsymbol{v}^0_\alpha, \operatorname{Var}(\hat{\boldsymbol{v}}^0_\alpha)\right)$$

where $\boldsymbol{v}^0_\alpha$ is defined in analogy to $\boldsymbol{\beta}^0_\alpha$ and

$$\begin{aligned}\operatorname{Var}(\hat{\boldsymbol{v}}^0_\alpha) = &\left(\sum_{i=1}^n \varpi^0_{i,\alpha}\boldsymbol{z}_i\boldsymbol{z}'_i + \boldsymbol{P}\right)^{-1} \\ &\left\{\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}'_i \operatorname{Var}\Big((w^0_{i,\alpha})\left(y_i - \boldsymbol{z}'_i\boldsymbol{v}^0_\alpha\right)\Big)\right\} \\ &\left(\sum_{i=1}^n \varpi^0_{i,\alpha}\boldsymbol{z}_i\boldsymbol{z}'_i + \boldsymbol{P}\right)^{-1}.\end{aligned} \qquad \text{(C.12)}$$

The covariance matrix of the penalized estimate has the typical sandwich form arising from the inclusion of the penalty in the estimation objective.

As before, the residual terms $y_i - \boldsymbol{z}'_i\boldsymbol{v}^0_\alpha$ in equation (C.12) can be replaced by empirical terms in order to estimate the variance, where in close analogy to (C.11) we divide the fitted version $\left(y_i - \boldsymbol{z}'_i\hat{\boldsymbol{v}}^0_\alpha\right)^2$ by its generalized hat matrix entry

$$1 - w^0_{i,\alpha}\boldsymbol{z}'_i\left(\sum_{j=1}^n w^0_{j,\alpha}\boldsymbol{z}_j\boldsymbol{z}'_j + \boldsymbol{P}\right)^{-1}\boldsymbol{z}_i. \qquad \text{(C.13)}$$

Of course, in practice the smoothing parameters will have to be determined jointly with the regression coefficients to obtain a data-driven amount of smoothness. A REML estimate based on the Schall algorithm (Schall, 1991) has been adapted to expectiles by Schnabel and Eilers (2009b). In our simulations and the example, we will use asymmetric cross-validation adapted for geoadditive expectile regression in Sobotka and Kneib (2012). The grid search over the smoothing parameter for the minimal cross-validation score is widened to an r-dimensional grid. The score itself is defined as

$$V_g^w = \frac{n\sum_{i=1}^n w_{i,\alpha}(y_i - \eta_{\alpha,i})^2}{[\mathrm{tr}(\mathbb{1} - \boldsymbol{H}^\alpha)]^2}$$

and the score is therefore independent from the number of functions r. The method is computationally demanding but accurate. We use the more accurate possibility to gain reliable informations on the confidence interval performance.

C.3 Confidence Intervals

C.3.1 Asymptotic Confidence Intervals

Equation (C.12) together with the correction (C.13) provides us with the asymptotic covariance matrix of the complete estimate $\hat{\boldsymbol{v}}_\alpha$ and therefore the covariance matrix of specific coefficient vectors of interest can immediately be obtained by extracting the appropriate sub-blocks. For example, for the variance of the estimated function evaluation $\hat{f}_{j,\alpha}(x_{i(\tilde{d}+j)}) = \boldsymbol{b}_{ij}'\boldsymbol{u}_{j,\alpha}$, we obtain

$$\mathrm{Var}(\hat{f}_{j,\alpha}(x_{i(\tilde{d}+j)})) = \boldsymbol{b}_{ij}' \,\mathrm{Var}(\hat{\boldsymbol{u}}_{j,\alpha})\boldsymbol{b}_{ij}$$

where $\mathrm{Var}(\hat{\boldsymbol{u}}_{j,\alpha})$ is the block of $\mathrm{Var}(\hat{\boldsymbol{v}}_\alpha)$ corresponding to $\hat{\boldsymbol{u}}_{j,\alpha}$. Together with the asymptotic normality of the least asymmetrically weighted squares estimate, this yields the following confidence interval for the true function evaluation $f_{j,\alpha}(x_{i(\tilde{d}+j)})$:

$$\mathrm{CI}(\hat{f}_{j,\alpha}(x_{i(\tilde{d}+j)})) = \left[\hat{f}_{j,\alpha}(x_{i(\tilde{d}+j)}) \pm z_{1-\frac{a}{2}}\sqrt{\mathrm{Var}(\hat{f}_{j,\alpha}(x_{i(\tilde{d}+j)}))}\right]$$

where $z_{1-\frac{a}{2}} = \Phi^{-1}(1-\frac{a}{2})$ is the $(1-\frac{a}{2})$-quantile of the standard normal distribution. Note that a particular amount of undercoverage is inevitable since we work with normal but not t-distribution quantiles.

C.3.2 Bootstrap Confidence Intervals

A further possibility to fit pointwise $(1-a)$-confidence intervals to expectile regression curves can be created with large computational expense. By conducting a nonparametric bootstrap, the distribution of the estimated expectiles can be approximated. At first, B bootstrap samples $(\boldsymbol{y}, \boldsymbol{X})^*_{b=1,\ldots,B}$ are drawn from the original data set. The expectiles are fitted independently for all B samples resulting in a bootstrapped sample $m_\alpha(\boldsymbol{x}^*_{i,1}), \ldots, m_\alpha(\boldsymbol{x}^*_{i,B})$ from the unknown distribution of the true expectile $m_\alpha(x_i)$. According to Efron and Tibshirani (1993) for a number of bootstrap replications $B \geq 1000$ we can construct bootstrap percentile intervals from $m_\alpha(\boldsymbol{x}^*_{i,1}), \ldots, m_\alpha(\boldsymbol{x}^*_{i,B})$ with sufficient quality. This holds under the assumption that the empirical distribution formed by the observations $(y_i, x_i)_{i=1,\ldots,n}$, is a good estimate for the unknown true distribution. The resulting pointwise intervals are therefore constructed from the $(\frac{a}{2}B)$-th and the $((1-\frac{a}{2})B)$-th element of the sorted set of the expectile estimates for each of the effects f_j from the Bootstrap samples and $i = 1, \ldots, n$:

$$\mathrm{CI}(\hat{f}^*_{j,\alpha}(x_i)) = \left[\hat{f}_{j,\alpha}(\boldsymbol{x}^*_{b_1,i})_{(\frac{a}{2}B)}\,;\,\hat{f}_{j,\alpha}(\boldsymbol{x}^*_{b_2,i})_{((1-\frac{a}{2})B)}\right].$$

An alternative would be to construct bootstrap-t-intervals. This would require an additional nonparametric bootstrap inside every previously drawn bootstrap sample to estimate the variance of the expectiles. In consequence this method would take a lot of time or processor cores when used on large data sets. Therefore we restrict our analyses to the bootstrap percentile intervals.

C.4 Empirical Evaluation

C.4.1 Simulation Study

After introducing two estimation approaches for expectile regression confidence intervals, an asymptotic and a numerical method, their merits and disadvantages will now be investigated in terms of a simulation study. The data structures considered in the simulation study are linear on the one hand, mixed and additive nonlinear on the other in order to simulate different data scenarios. We will also investigate numerical properties of the estimation approaches in terms of computing time.

Design

The models used for the simulations are defined as

$$y = 0.75 + 0.9x_1 + \epsilon \tag{C.14}$$

$$y = 3x_3 + 3\underbrace{\exp(-x_1^2)}_{f_{\text{p-spline}}(x_1)} + \epsilon \tag{C.15}$$

$$y = \underbrace{x_1^2}_{f_{\text{p-spline}}(x_1)} + \underbrace{\sin(8x_2 - 4) + 2\exp(-(16x_2 - 8)^2)}_{f_{\text{p-spline}}(x_2)} + \epsilon \tag{C.16}$$

where ϵ follows either a normal distribution $N(0, 3^2)$, a beta distribution or the so called "expectiles-meet-quantiles" (emq) distribution with distribution function

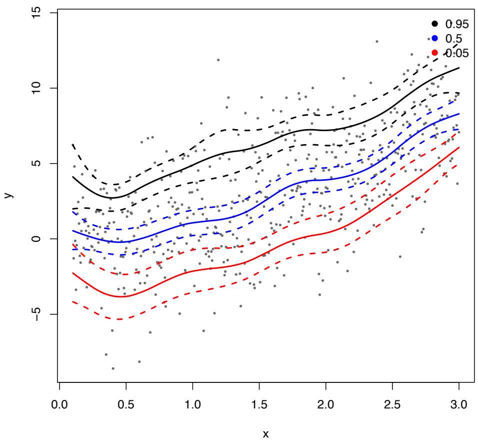
$$F_{\mu,s}(\epsilon) = 0.5\left(1 + \text{sign}(\epsilon - \mu)\sqrt{1 - \frac{2}{2 + \left(\frac{\epsilon-\mu}{s}\right)^2}}\right)$$

with expectation $\mu = 0$ and scaling parameter $s = \sqrt{2}$ (The variance itself is not finite regardless the value of s). The latter distribution has the desirable property that quantiles and expectiles coincide (see Koenker, 2005, p. 67) for all parameters $\mu \in \mathbb{R}$ and $s > 0$. Also, due to the non-existing second moments a key assumption for the asymptotic results is violated. Here, we can examine the importance of the assumption. Note that both the normal and the "emq" distribution are homoscedastic while the beta distribution is variance heteroscedastic with Beta$(0.5x_1, 3x_1)$ for models (C.14) and (C.15) and Beta$(0.5x_1, 3x_2)$ for model (C.16). The true expectiles of the above distributions are obtained by numerically solving

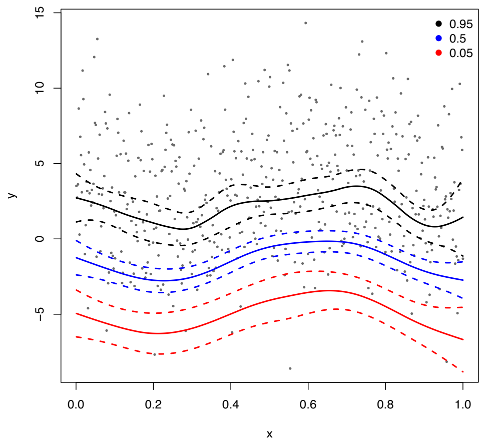
$$\alpha = \frac{G(m_\alpha) - m_\alpha F(m_\alpha)}{2(G(m_\alpha) - m_\alpha F(m_\alpha)) + (m_\alpha - m_{0.5})}$$

where F is the cumulative distribution function, $G(m_\alpha) = \int_{-\infty}^{m_\alpha} y \; dF(y)$ is the partial moment function and $G(\infty) = m_{0.5}$ is the expectation of y.

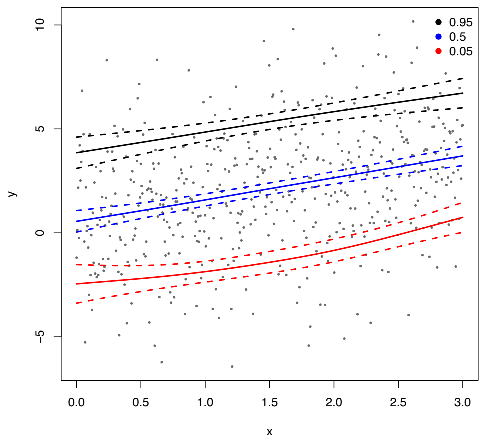
The binary covariate x_3 is drawn from a $B(1, 0.5)$ distribution. The values of the continuous covariates x_1 and x_2 are equally spaced over their domains $[0; 3]$ and $[0; 1]$, respectively. Therefore we have the same positions in every simulated data set where the confidence intervals are evaluated. The corresponding functions are modeled as cubic penalized splines with second order difference penalty and 20 inner knots.



(a) Nonlinear effect $f_{\text{p-spline}}(x_1)$



(b) Nonlinear effect $f_{\text{p-spline}}(x_2)$



(c) Linear effect of x_1

Figure C.1: Exemplary data and fitted asymptotic confidence intervals for one simulated data set with $n = 500$ observations and $N(0, 3^2)$ distributed errors.

Figure C.1 visualizes simulated data for one replication to give an impression of the functional form of the effects considered. Based on sample sizes of $n = 100$, 250, 500 and 1000, we generated 1000 simulation replications for each of the 36 different data structures arising from the combination of (i) the model (linear, mixed and additive), (ii) the error distribution (normal, beta, emq), and (iii) the sample size. For each data set, we applied the two different approaches for the estimation of confidence intervals introduced in the previous section, i.e. asymptotic normality and bootstrap percentiles to determine confidence intervals for expectiles with asymmetries $\alpha \in \{0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 0.8, 0.9, 0.95, 0.98, 0.99\}$. The asymptotic normality was used to estimate confidence intervals from the regression coefficients obtained from least asymmetrically weighted squares (LAWS). The same is true for the bootstrap percentile intervals.

All simulations have been implemented using `expectreg` (see Sobotka, Schnabel, and Schulze Waltrup, 2013), a package for R (R Development Core Team, 2010). The package also contains expectile functions for several distributions including those used in the simulations.

Performance measures

For the measurement of the quality of the results we evaluate the true expectile curve at the covariate values $x_{1,i}$ and $x_{2,i}$. Then the number of times the true expectile is covered by the interval are counted. Also the intervals will be compared according to their width. We therefore measure the coverage of the confidence intervals for $h = 1, 2$ at a given covariate value $x_{h,i}$ as

$$\widehat{\text{Cover}}(CI(\hat{f}_{j,\alpha}(x_{h,i})) = \frac{1}{1000} \sum_{k=1}^{1000} \mathbb{1}_{\{\hat{f}_{j,\alpha}(x_{h,i}) \in CI(\hat{f}^{[k]}_{j,\alpha}(x_{h,i}))\}},$$

the maximum width of all confidence intervals at all fixed $x_{h,i}$

$$\max \widehat{\text{Width}}(CI(\hat{f}_{j,\alpha}(x_{h,i}))) = \max_k(\hat{f}^{[k]}_{j,\alpha,U}(x_{h,i}) - \hat{f}^{[k]}_{j,\alpha,L}(x_{h,i}))$$

and for a compact measure the mean coverage along the covariate x_h, $h = 1, 2$

$$\overline{\widehat{\text{Cover}}}(CI(\hat{f}_{j,\alpha}(x_h))) = \frac{1}{1000n} \sum_{i=1}^{n} \sum_{k=1}^{1000} \mathbb{1}_{\{\hat{f}_{j,\alpha}(x_{h,i}) \in CI(\hat{f}^{[k]}_{j,\alpha}(x_{h,i}))\}} \tag{C.17}$$

as well as the mean interval width

$$\widehat{\overline{\text{Width}}}(CI(\hat{f}_{j,\alpha}(x_h))) = \frac{1}{1000n} \sum_{i=1}^{n} \sum_{k=1}^{1000} \hat{f}^{[k]}_{j,\alpha,U}(x_{h,i}) - \hat{f}^{[k]}_{j,\alpha,L}(x_{h,i})\,. \tag{C.18}$$

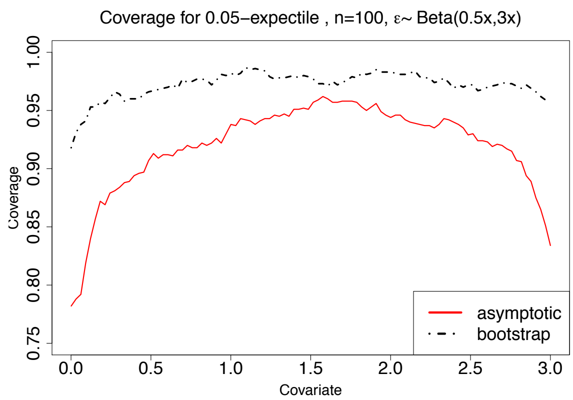
Here, $\hat{f}^{[k]}_{j,\alpha}$ denotes the expectile estimate for the j-th effect in the k-th simulation run. Further, the upper or lower end of the interval is indicated by U and L, respectively. In order to get a better hold of the actual quality of the confidence intervals we guarantee identifiability of the expectiles in the additive model by centering $\tilde{f}_{j,\alpha}(x_i) = f_{j,\alpha}(x_i) - \bar{f}_{j,\alpha}$.

Results

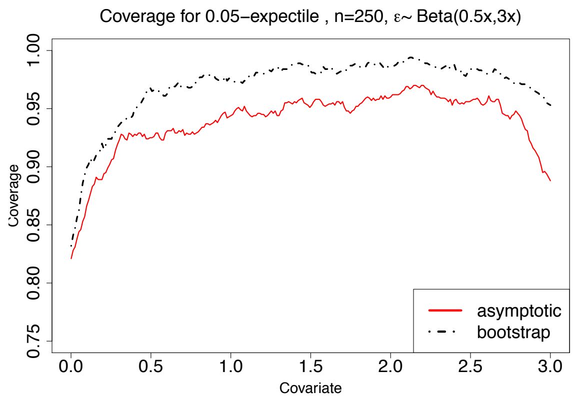
The first observation we can make is that the desired confidence level of 95% cannot be guaranteed for all situations. None of the introduced methods shows that quality. The best results are achieved for the special case of a mean regression ($\alpha = 0.5$) and for covariate values near $\bar{x}$. The larger the asymmetry ($\alpha \to 0$ or $\alpha \to 1$) and the nearer to the edge of the covariates' support, the higher the probability that the confidence level will not be met. The former is displayed in Table C.1, the latter is exemplary shown for the beta distribution in Figure C.2. In addition, calculating the mean coverage for all covariates, as defined in Section C.4.1, results in the simulated coverage probabilities shown in Table C.1. Results for $n = 250$ and the width of the confidence intervals are available on request.

We also investigate the average width of the confidence intervals. Apparently, for symmetrical distributions the width increases towards the boundary of the covariate support. For the heteroscedastic scenario this needs not to be the case as the beta distribution shows. Table C.1 shows an increasing coverage probability with growing sample size. The latter, however, is only partly true for the emq distribution due to the infinite variance. In comparison, the gain in coverage probability and the decrease in interval width is stronger for the confidence intervals constructed from the asymptotic properties. The latter is especially important since we want the narrowest interval width possible given a proper coverage. Analysing both measures together, the coverage (C.17) and the width (C.18), ensures that we select intervals for which the appropriate coverage is not gained by additional interval width.

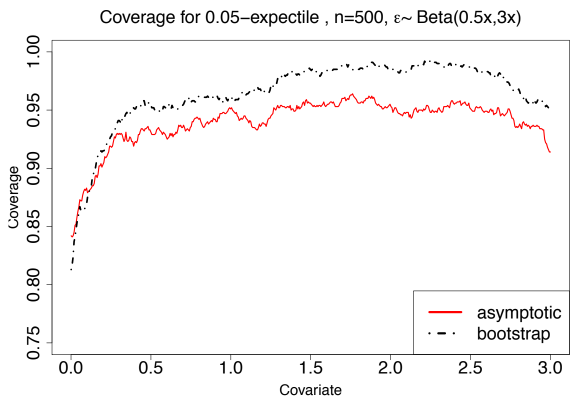
Regarding the performance of the bootstrap percentile intervals one needs to bear in mind the increased computational demand. In fact, one needs to fit the complete set of considered expectiles in each nonparametric bootstrap samples which is a rather time-consuming method. After this computational burden, that can take more than an hour for a single data set, depending on the complexity of the data, the results are however satisfactory. The



(a) relative coverage for $n = 100$



(b) relative coverage for $n = 250$



(c) relative coverage for $n = 500$

Figure C.2: Simulation results for $\alpha = 0.05$ with $n = 100, 250, 500$ observations and Beta$(0.5x, 3x)$ distributed errors. The relative coverage frequency for both methods along the covariate is shown. The method of asymptotic normality is plotted in red, the LAWS bootstrap percentile intervals in black and dotted.

$n = 100$						
error	$F_{emq}(0, \sqrt{2})$		$N(0, 3^2)$		$\text{Beta}(0.5x_1, 3x_1)$	
α	boot	asympt	boot	asympt	boot	asympt
0.01	0.358	0.345	0.706	0.715	0.947	0.858
0.02	0.496	0.462	0.783	0.788	0.950	0.880
0.05	0.654	0.615	0.848	0.847	0.952	0.899
0.1	0.758	0.728	0.882	0.875	0.949	0.902
0.2	0.841	0.829	0.901	0.893	0.941	0.911
0.5	0.914	0.939	0.920	0.915	0.918	0.903
0.8	0.824	0.819	0.910	0.899	0.881	0.858
0.9	0.733	0.714	0.886	0.874	0.848	0.829
0.95	0.623	0.600	0.850	0.841	0.813	0.797
0.98	0.467	0.439	0.776	0.780	0.760	0.725
0.99	0.332	0.319	0.700	0.723	0.711	0.657
$n = 500$						
error	$F_{emq}(0, \sqrt{2})$		$N(0, 3^2)$		$\text{Beta}(0.5x_1, 3x_1)$	
α	boot	asympt	boot	asympt	boot	asympt
0.01	0.634	0.594	0.876	0.851	0.946	0.923
0.02	0.713	0.681	0.902	0.879	0.946	0.927
0.05	0.801	0.771	0.922	0.905	0.930	0.928
0.1	0.848	0.830	0.929	0.915	0.921	0.925
0.2	0.893	0.890	0.930	0.924	0.924	0.922
0.5	0.922	0.947	0.929	0.931	0.934	0.931
0.8	0.882	0.873	0.931	0.930	0.895	0.900
0.9	0.833	0.814	0.931	0.923	0.843	0.874
0.95	0.780	0.752	0.923	0.913	0.783	0.862
0.98	0.696	0.661	0.904	0.888	0.721	0.846
0.99	0.604	0.573	0.882	0.855	0.687	0.829
$n = 1000$						
error	$F_{emq}(0, \sqrt{2})$		$N(0, 3^2)$		$\text{Beta}(0.5x_1, 3x_1)$	
α	boot	asympt	boot	asympt	boot	asympt
0.01	0.714	0.665	0.937	0.879	0.929	0.934
0.02	0.790	0.734	0.939	0.905	0.933	0.937
0.05	0.858	0.802	0.932	0.919	0.921	0.939
0.1	0.897	0.843	0.932	0.926	0.911	0.934
0.2	0.919	0.890	0.936	0.932	0.923	0.927
0.5	0.925	0.942	0.946	0.935	0.947	0.934
0.8	0.889	0.888	0.942	0.931	0.912	0.910
0.9	0.876	0.846	0.933	0.926	0.824	0.894
0.95	0.856	0.808	0.930	0.915	0.697	0.888
0.98	0.783	0.741	0.913	0.898	0.603	0.873
0.99	0.708	0.681	0.884	0.877	0.563	0.861

Table C.1: Mean relative coverage frequency as defined in Equation (C.17) for the eleven asymmetry parameters, both estimation methods and all error distributions for a sample size of $n = 100, 500, 1000$.

bootstrap intervals provide a coverage of nearly $1 - a$ with the limitations stated in the beginning. Especially for small samples, the provided coverage of the bootstrap method is better than from the asymptotic method without resulting in unreasonably wide intervals. Also for small samples the time required to conduct the bootstrap is within a few minutes depending on the possibilities for parallelization.

In conclusion, we can see that both methods have their merits and weaknesses. Small samples are best tackled with bootstrap intervals and for heteroscedastic errors or large samples we can recommend to use the asymptotic normality to construct confidence intervals for the expectile curves. If the variance does not exist, we can see that the violated assumption in the asymptotics leads to poor coverage. In simple cases, 500 observations will suffice. Otherwise and if extreme expectiles like $\alpha = 0.01, 0.99$ shall be estimated, more are required.

C.4.2 Childhood Malnutrition in India

Malnutrition is a severe problem in developing countries. Regular surveys are therefore conducted on national bases in order to determine risk factors for malnutrition. General and representative studies on health and population development are done by MEASURE Demographic and Health Surveys (DHS). Those include topics like HIV distribution, fertility or nutrition aspects. The data can be obtained from `www.measuredhs.com` free of charge for research purposes. In our case we use data on childhood malnutrition in India from the year 2001. After preprocessing and deleting observations with missing values, the data contains 24316 observations in 40 variables. In general, malnutrition of each individual i is measured as a score Z defined as

$$Z_i = \frac{AC_i - M}{s}$$

where AC is an anthropometric characteristic. Most of the time the weight in relation to the age is measured for this variable. This characteristic is standardized by subtracting the median M and dividing by the standard deviation s of the same attribute in a reference population. While a score based on weight is also an indicator for acute malnutrition, an insufficient height for a child's age, also called stunting, is a distinct indicator for chronic malnutrition. Therefore stunting is the variable that is modeled here. The score for stunting Z is neither normally distributed nor restricted to a certain support. In our data the value ranges from -600 to 600. The model is inspired by Fenske, Kneib, and Hothorn (2011) and the predicted stunting η_α for the α-expectile is modeled as

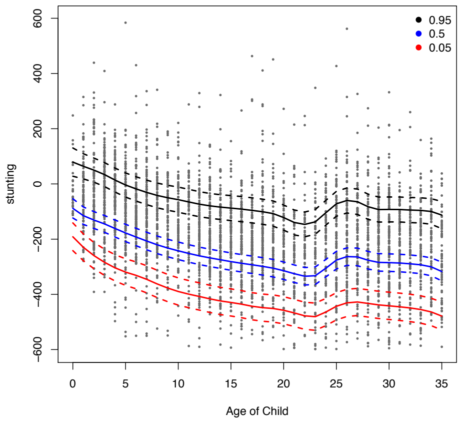
$$\begin{aligned}\eta_\alpha =& \boldsymbol{x}'\boldsymbol{\beta}_\alpha + f_{\alpha,1}(\text{age of child}) + f_{\alpha,2}(\text{duration of breastfeeding})\\ &+ f_{\alpha,3}(\text{BMI of mother}) + f_{\alpha,4}(\text{age of mother})\\ &+ f_{\alpha,5}(\text{education years of mother})\\ &+ f_{\alpha,6}(\text{education years of partner}) + f_{\alpha,\text{spat}}(\text{district}).\end{aligned}$$

The parametric effects included in $\boldsymbol{x}$ are listed in Table C.2. Further there are six nonlinear effects in the model that are fitted with a cubic P-spline basis constructed from 20 inner knots and penalized with second order differences. Also one spatial effect is included as a Markov random field. A special interest of this analysis lies in the lower tails of the conditional distribution of Z. The expectiles for small values of α will show the relation of the covariates to the response for cases of severe malnutrition. Confidence intervals from a nonparametric bootstrap cannot be considered here as we expect a computing time of several weeks.

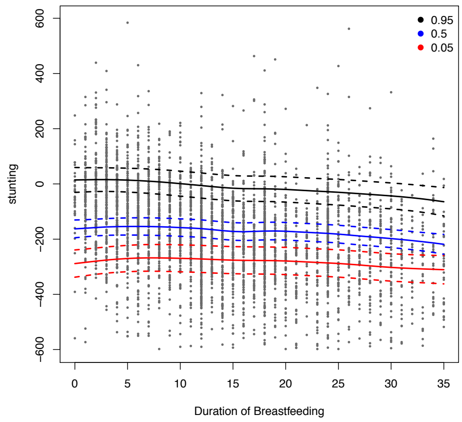
For the lower expectiles we can see that stunting gets worse if the child is later in the birth order. This as well as the insignificance of the residence region of the mother (rural / urban) is a result comparable to the lower quantiles computed by Fenske, Kneib, and Hothorn (2011). The 0.8 and 0.95-expectiles show a different behaviour for these covariates. The family size is insignificant for children that do not suffer from stunting. For those children living in urban areas it also has a positive effect. We can support this by the 0.95-expectile of the regions of India depicted in Figure C.4. The map shows a positive effect on the nutritional status of the children for densely populated areas. Those regions are mainly in the northeast along the rivers Ganges and Brahmaputra. In consequence, we can assume a sufficient supply with fresh water for these children. We can also see a relation to the effects of the religion here since most of India's Muslims live in the densely populated areas. The inclusion of an interaction term could be part of further research. In the additive model considered here, an increased correlation between two covariates will just result in wider confidence intervals. The effects nevertheless show us that Muslim children suffer from stunting less than children from the other religions. This observation can be made throughout all expectiles and stands in contrast to the results from Fenske, Kneib, and Hothorn (2011) whose results indicated no difference between the five religions. This might be due to the fact that no spatial effect could be included in the quantile regression model. They also performed variable selection in the quantile regression which led to the elimination of the television indicator variable from their model. The expectiles, however, show that the presence of a TV in a household is an indicator for less stunting. The reason for this result is probably that the whole family

Variable / α	0.05	0.2	0.8	0.95
sex of child	-2.91	-2.45	-1.35	3.52
reference: "male"	*(-8.63, 2.80)*	*(-7.31, 2.41)*	*(-6.46, 3.74)*	*(-2.88, 9.94)*
twin birth	**-67.53**	**-68.71**	**-72.21**	**-79.22**
reference: "single birth"	*(-91.01, -44.10)*	*(-88.25, -49.17)*	*(-93.84, -50.59)*	*(-112.10, -46.34)*
birth order:		*reference: "first"*		
"second"	-5.75	**-8.81**	**-7.66**	0.05
	(-13.37, 1.87)	*(-15.57, -2.06)*	*(-14.71, -0.61)*	*(-9.03, 9.13)*
" third"	**-15.70**	**-15.82**	**-14.55**	-7.45
	(-25.28, -6.11)	*(-23.82, -7.82)*	*(-23.18, -5.92)*	*(-19.30, 4.38)*
"fourth"	**-17.97**	**-17.25**	-4.07	**18.35**
	(-29.12, -6.81)	*(-26.64, -7.86)*	*(-13.90, 5.74)*	*(3.65, 33.05)*
"fifth"	**-35.54**	**-33.41**	**-24.00**	-9.91
	(-47.59, -23.49)	*(-43.11, -23.71)*	*(-34.18, -13.81)*	*(-25.56, 5.72)*
mother's work	-1.41	-3.48	-1.25	2.20
reference: "unemployed"	*(-7.81, 4.97)*	*(-9.33, 2.36)*	*(-7.46, 4.95)*	*(-6.37, 10.79)*
mother's religion		*reference: "christian"*		
"hindu"	**-7.96**	-4.91	-2.39	1.05
	(-15.47, -0.46)	*(-10.55, 0.72)*	*(-8.62, 3.83)*	*(-9.44, 11.55)*
"muslim"	**31.23**	**24.27**	**26.51**	**37.91**
	(19.14, 43.32)	*(13.01, 35.53)*	*(14.44, 38.59)*	*(22.16, 53.66)*
"sikh"	6.72	5.27	8.51	8.23
	(-16.30, 29.75)	*(-11.46, 22.00)*	*(-7.79, 24.82)*	*(-15.94, 32.41)*
"other"	**22.82**	**14.49**	7.77	5.41
	(6.24, 39.41)	*(0.37, 28.61)*	*(-7.15, 22.69)*	*(-15.40, 26.23)*
mother's residence	-1.61	-0.79	2.32	**8.98**
reference: "rural"	*(-8.34, 5.11)*	*(-5.64, 4.04)*	*(-2.44, 7.09)*	*(1.99, 15.97)*
# dead children:		*reference: "0"*		
"1"	-5.62	-2.89	-6.18	-10.43
	(-12.94, 1.69)	*(-8.23, 2.45)*	*(-13.05, 0.68)*	*(-21.28, 0.42)*
"2"	-3.80	-1.46	-6.05	-11.91
	(-16.85, 9.24)	*(-12.21, 9.28)*	*(-18.92, 6.80)*	*(-32.28, 8.45)*
"3"	-15.94	**-16.05**	-14.93	-16.26
	(-33.82, 1.92)	*(-31.88, -0.23)*	*(-35.07, 5.20)*	*(-44.51, 11.99)*
electricity supply	**16.71**	**12.65**	7.73	4.77
reference: "no"	*(9.47, 23.95)*	*(5.80, 19.50)*	*(-0.35, 15.82)*	*(-6.18, 15.72)*
radio	3.47	**4.51**	**5.58**	3.52
reference: "no"	*(-1.80, 8.75)*	*(0.72, 8.31)*	*(1.44, 9.73)*	*(-3.73, 10.78)*
television	**11.49**	**13.35**	**14.80**	**18.77**
reference: "no"	*(4.73, 18.25)*	*(8.57, 18.13)*	*(9.61, 19.99)*	*(10.50, 27.04)*
refrigerator	9.73	**10.63**	**9.29**	**6.93**
reference: "no"	*(-1.18, 20.65)*	*(2.17, 19.09)*	*(0.20, 18.37)*	*(10.50, 27.04)*
bicycle	-3.87	-3.53	**-7.36**	**-8.83**
reference: "no"	*(-8.80, 1.05)*	*(-7.35, 0.28)*	*(-12.60, -2.12)*	*(-16.66, -1.01)*
motorcycle	**11.13**	**9.80**	**9.56**	**11.61**
reference: "no"	*(2.26, 20.00)*	*(2.88, 16.72)*	*(1.91, 17.22)*	*(0.31, 22.90)*
car	-10.55	1.10	4.40	11.09
reference: "no"	*(-38.11, 17.00)*	*(-12.48, 14.69)*	*(-9.45, 18.26)*	*(-14.19, 36.38)*

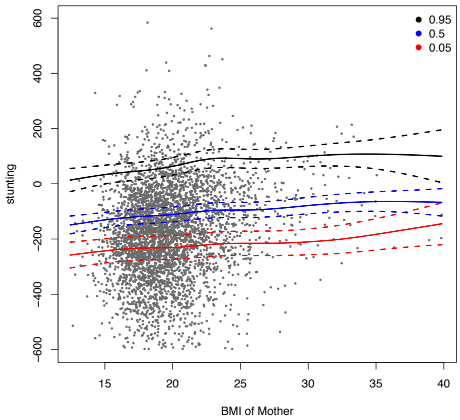
Table C.2: Estimated parametric effects for Childhood Malnutrition data. Reference categories and confidence intervals ($\alpha = 0.95$) obtained by asymptotic normality are included in italics. Significant effects are set in bold.



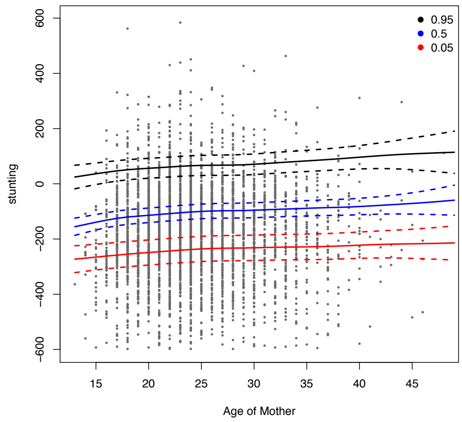
(a) age of child



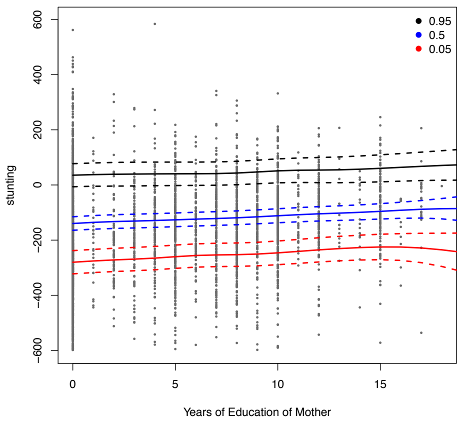
(b) duration of breastfeeding



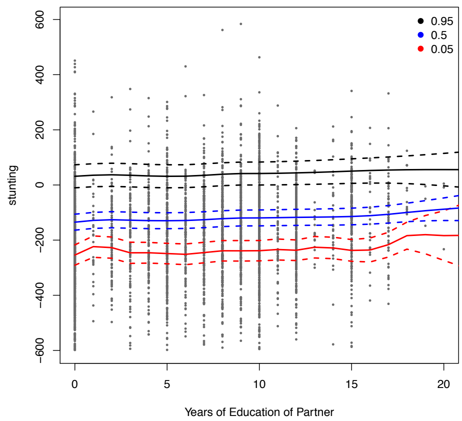
(c) BMI of mother



(d) age of mother

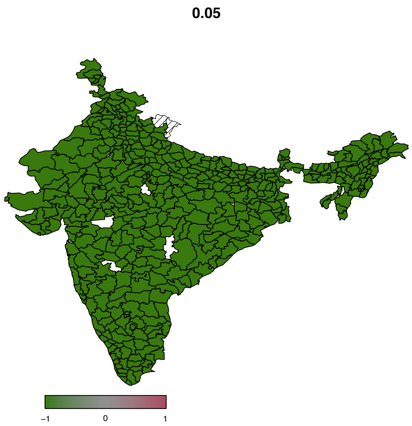


(e) years of mother's education

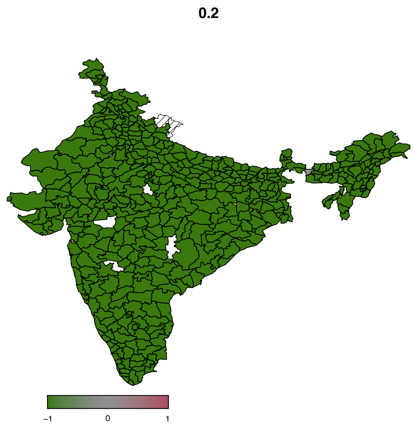


(f) partner's education

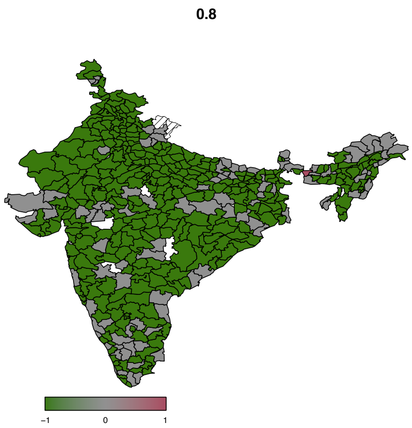
Figure C.3: Estimated nonlinear effects and confidence intervals for the six continuous covariates included in the model for the 0.05, 0.5 and 0.95-expectile.



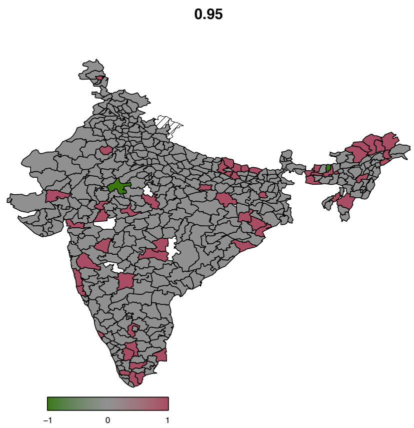
(a) 0.05-expectile



(b) 0.2-expectile



(c) 0.8-expectile



(d) 0.95-expectile

Figure C.4: Estimated significance indicators for the effects of the Markov random field in the regions of India for four expectiles. Green regions indicate a significant negative effect on the response while pink regions indicate a positive effect.

is provided with food before the money is spent on a TV. So we can take this variable as an indicator for wealth. Not yet mentioned was the positive influence of the presence of a motorcycle or a refrigerator to the stunting score.

From the six continuous covariates included in the model and shown in Figure C.3 we see that up to an age of two years the stunting gets worse and after that there's a consolidation. The remaining five continuous effects present less drastic changes along the covariates than the quantiles portrayed. For increasing age, BMI and years of education of the mother we observe a slight increase in the stunting score. Comparing both the education of the mother and of her partner we make the same observation as Fenske, Kneib, and Hothorn (2011). The education of the partner is less important for the nutritional status of the child. For all continuous variables we can see a homoscedastic behaviour as the expectiles are almost parallel throughout the support of the covariates. Also we can conclude from the expectiles that the conditional distribution of the stunting score is right skewed. Further, the variation in the response is substantial. This leads to wide confidence intervals to all expectiles even with the large amount of observations. The latter is nevertheless important for the high smoothness of the expectile curves. The analyses demonstrate several indicators that are associated with malnutrition in India. But we can see from the lower expectiles in Figure C.4, severe malnutrition can be found anywhere in India.

C.5 Conclusion

In this paper, we derived the asymptotic results supplementing the point estimators for geoadditive expectiles. The asymptotic normality of the LAWS method as well as the subsequent confidence intervals are an essential extension to the estimation methods introduced e.g. in Sobotka and Kneib (2012). Our simulations and the application to the malnutrition data have shown us that we can safely replace the computationally expensive method of the bootstrap with the usage of the asymptotic properties. As Figure C.2 has shown, both methods provide similar coverage for growing sample sizes.

Generally, we need to recollect that the advantages of expectile regression over mean regression can be exploited solely when regarding a set of expectiles. As seen in Section C.4.2, by comparing different expectiles we gain information about the distribution of the response. The introduced confidence intervals help us by signifying the strength of the results. The data analysis has also shown that the obtained information is comparable to a quantile regression despite reduced interpretability. Hence, we use expectiles and gain computational advantages and flexible geoadditive models.

Acknowledgements

We thank two anonymous referees and a coordinating editor for valuable comments that lead to considerable improvement upon the original version of this paper. Financial support from the German Research Foundation (DFG) grant KN 922/4-1 is gratefully acknowledged.

Proof of Asymptotic Normality

Proof. Note first that

$$\begin{aligned}
&\sum_{i=1}^{n} w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha^0)\boldsymbol{x}_i(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0) \\
&= \sum_{i=1}^{n} w_{i,\alpha}(\boldsymbol{\beta}_\alpha^0)\boldsymbol{x}_i(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0) \\
&+ \sum_{i=1}^{n}\big(w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha^0) - w_{i,\alpha}(\boldsymbol{\beta}_\alpha^0)\big)\boldsymbol{x}_i(y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0)
\end{aligned} \tag{C.19}$$

with

$$w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha^0) - w_{i,\alpha}(\boldsymbol{\beta}_\alpha^0) = \begin{cases}
0, & \text{for } y_i \geq \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0 \\
 & \text{and } y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0 \\
\alpha - (1-\alpha), & \text{for } y_i \geq \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0 \\
 & \text{and } y_i < \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0 \\
(1-\alpha) - \alpha, & \text{for } y_i < \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0 \\
 & \text{and } y_i \geq \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0 \\
0, & \text{for } y_i < \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_\alpha^0 \\
 & \text{and } y_i < \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0.
\end{cases}$$

Since $\hat{\boldsymbol{\beta}}_\alpha^0 - \boldsymbol{\beta}_\alpha^0 = O_p(n^{-1/2})$, the last component in (C.19) is of ignorable asymptotic order $O_p(1)$ (while the other component is of order $O_p(n^{1/2})$). Following the same line of arguments, we can now derive the asymptotic properties for the final estimate

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_\alpha^0 - \boldsymbol{\beta}_\alpha^0 &= \left(\sum_{i=1}^{n} w_{i,\alpha}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\sum_{i=1}^{n} w_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha)\boldsymbol{x}_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0)\right) \\
&= \left(\sum_{i=1}^{n} \varpi_{i,\alpha}^0\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\sum_{i=1}^{n} \varpi_{i,\alpha}^0\boldsymbol{x}_i(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\alpha^0)\right) + O_p(n^{-1})
\end{aligned}$$

and therefore $\hat{\boldsymbol{\beta}}_\alpha \overset{a}{\sim} N\left(\boldsymbol{\beta}_\alpha^0, \mathrm{Var}(\hat{\boldsymbol{\beta}}_\alpha^0)\right)$. □

References

Abdous, B. and B. Remillard (1995). Relating quantiles and expectiles under weighted symmetry. *Annals of the Institute of Statistical Mathematics 47*(2), 371–384.

Aigner, D. J., T. Amemiya, and D. J. Poirier (1976). On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review 17*(2), 377–396.

Bollaerts, K., P. H. C. Eilers, and M. Aerts (2006). Quantile regression with monotonicity restrictions using P-splines and the L1-norm. *Statistical Modelling 6*(3), 189–207.

Bondell, H., B. Reich, and H. Wang (2010). Non-crossing quantile regression curve estimation. *Biometrika 97*(4), 825–838.

Breckling, J. and R. Chambers (1988). M-quantiles. *Biometrika 75*(4), 761–771.

Buchinsky, M. (1998). Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources 33*, 88–126.

Bühlmann, P. and T. Hothorn (2007, 11). Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science 22*(4), 477–505.

Chambers, R. L. and R. Dunstan (1986). Estimating distribution functions from survey data. *Biometrika 73*(3), 597–604.

Chen, Q., M. R. Elliott, and R. J. A. Little (2010). Bayesian penalized spline model-based inference for finite population proportion in unequal probability sampling. *Survey Methodology 36*(1), 23–34.

Chen, Q., M. R. Elliott, and R. J. A. Little (2012). Bayesian inference for finite population from unequal probability sampling. *Survey Methodology 38*(2), 203–214.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica 78*(3), 1093–1125.

Croissant, Y. and G. Millo (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software 27*(2).

Currie, I. D., M. Durbán, and P. H. C. Eilers (2006). Generalized array models with application to multidimensional smoothing. *Journal of the Royal Statistical Society B 68*(2), 259–280.

de Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.

De Rossi, G. and A. Harvey (2009). Quantiles, expectiles and splines. *Journal of Econometrics 152*(2), 179–185.

Diggle, P. J., K.-Y. Liang, and S. L. Zeger (1994). *Analysis of longitudinal data*. Oxford: Oxford University Press.

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

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. London: Chapman and Hall.

Eilers, P. H. C., I. D. Currie, and M. Durbán (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis 50*, 61–76.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Stat. Science 11*(2), 89–121.

Fahrmeir, L., T. Kneib, and S. Lang (2004). Penalized additive regression for space-time data: a Bayesian perspective. *Statistica Sinica 14*, 731–761.

Fahrmeir, L., T. Kneib, S. Lang, and B. Marx (2013). *Regression : Models, Methods and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Fenske, N., T. Kneib, and T. Hothorn (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association 106*(494), 494–510.

Godambe, V. P. (1991). Confidence intervals for quantiles. In V. P. Godambe (Ed.), *Estimating functions*, pp. 211–217. Oxford University Press.

Guo, M. and W. Härdle (2013). Simultaneous confidence bands for expectile functions. *AStA - Advances in Statistical Analysis 96*(4), 517–541.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*. London: Chapman and Hall.

He, X. (1997). Quantile curves without crossing. *The American Statistician 51*(2), 186–192.

Hofner, B., T. Hothorn, T. Kneib, and M. Schmid (2011). A framework for unbiased model selection based on boosting. *Journal of Computational and Graphical Statistics 20*, 956–971.

Hofner, B., A. Mayr, N. Fenske, and M. Schmid (2011). *gamboostLSS: Boosting Methods for GAMLSS Models*. R package version 1.0-3.

Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner (2013). *mboost: Model-Based Boosting*. R package version 2.2-2.

Jones, M. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics 44*(4), 721–727.

Jones, M. (1994). Expectiles and M-quantiles are quantiles. *Statistics & Probability Letters 20*(2), 149–153.

Kammann, E. E. and M. P. Wand (2003). Geoadditive models. *Applied Statistics 52*(1), 1–18.

Kauermann, G., T. Krivobokova, and L. Fahrmeir (2009). Some asymptotic results on generalized penalized spline smoothing. *Journal of the Royal Statistical Society, Series B 71*, 487–503.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kneib, T. (2013). Beyond mean regression (with discussion and rejoinder). *Statistical Modelling 13*(4), 275–385.

Kneib, T., F. Heinzl, A. Brezger, and D. Sabanés Bové (2013). *BayesX: R Utilities Accompanying the Software Package BayesX*. R package version 0.2-6.

Kocherginsky, M., X. He, and Y. Mu (2005). Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics 14*, 41–55.

Koenker, R. (1984). A note on L-estimates for linear models. *Statistics & Probability Letters 2*(6), 323–325.

Koenker, R. (1993). When are expectiles percentiles? *Econometric Theory 9*(3), 526–527.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis, Elsevier 91*(1), 74–89.

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

Koenker, R. (2013a). Discussion of "Beyond mean regression" by T. Kneib. *Statistical Modelling 13*(4), 323–333.

Koenker, R. (2013b). *quantreg: Quantile Regression*. R package version 4.97.

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

Koenker, R., P. Ng, and S. Portnoy (1994). Quantile smoothing splines. *Biometrika 81*(4), 673–680.

Koenker, R. and P. T. Ng (2005). Inequality constrained quantile regression. *Sankhya: The Indian Journal of Statistics 67*(2), 418–440.

Kuan, C. M., J. H. Yeh, and Y. C. Hsu (2009). Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics 150*(2), 261–270.

Kuk, A. Y. C. (1988). Estimation of distribution functions and medians under sampling with unequal probabilities. *Biometrika 75*(1), 97–103.

Lahiri, D. B. (1951). A method of sample selection providing unbiased ratio estimates. *Bulletin of the International Statistical Institute* (33), 133–140.

Leng, C. and W. Zhang (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing 24*(1), 123–136.

Leorato, S., F. Peracchi, and A. V. Tanase (2012). Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics & Data Analysis 56*(4), 768 – 784.

Mayr, A., N. Fenske, B. Hofner, T. Kneib, and M. Schmid (2012). Gamlss for high-dimensional data - a flexible approach based on boosting. *Applied Statistics 61*, 403–427.

Meinshausen, N. and P. Bühlmann (2010). Stability selection (with discussion). *Journal of the Royal Statistical Society, Series B 72*(4), 417–473.

Muggeo, V., M. Sciandra, A. Tomasello, and S. Calvo (2013). Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environmental and Ecological Statistics 20*, 519–531.

Murthy, M. N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

Neocleous, T. and S. Portnoy (2008). On monotonicity of regression quantile functions. *Statistics & Probability Letters 78*(10), 1226–1229.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing. *Econometrica 55*(4), 819–847.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science 1*(4), 502–518.

O'Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation. *SIAM J. Sci. Statist. Comput. 9*, 531–542.

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

R Core Team (2013). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rao, J. N. K., J. G. Kovar, and H. J. Mantel (1990). On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika 77*(2), 365–375.

Reiss, P. and L. Huang (2012). Smoothness selection for penalized quantile regression splines. *International Journal of Biostatistics 8*(1). Article 10.

Rigby, R. A. and D. M. Stasinopoulos (2005). Generalized additive models for location, scale and shape (with discussion). *Applied Statistics 54*(3), 507–554.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields.* Boca Raton: Chapman and Hall.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression.* Cambridge: Cambridge University Press.

Ruppert, D., M. P. Wand, and R. J. Carroll (2009). Semiparametric regression during 2003-2007. *Electronic Journal of Statistics 3*, 1193–1256.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*(4), 719–727.

Schnabel, S. K. and P. H. C. Eilers (2009a). An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research 21*(5), 109–134.

Schnabel, S. K. and P. H. C. Eilers (2009b). Optimal expectile smoothing. *Computational Statistics & Data Analysis 53*(12), 4168–4177.

Schnabel, S. K. and P. H. C. Eilers (2013a). A location scale model for non-crossing expectile curves. *Stat : the ISI's Journal for the Rapid Dissemination of Statistics Research 2*(1), 171 – 183.

Schnabel, S. K. and P. H. C. Eilers (2013b). Simultaneous estimation of quantile curves using quantile sheets. *Advances in Statistical Analysis 97*(1), 77–87.

Schnabel, S. K. and P. H. C. Eilers (2014). A note on expectile sheets. in revision.

Schulze Waltrup, L., F. Sobotka, T. Kneib, and G. Kauermann (2014). Expectile and quantile regression - David and Goliath? *Statistical Modelling*. to appear.

Sobotka, F., G. Kauermann, L. Schulze Waltrup, and T. Kneib (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing 23*(2), 135–148.

Sobotka, F. and T. Kneib (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis 56*(4), 755–767.

Sobotka, F., S. Schnabel, and L. Schulze Waltrup (2013). *expectreg: Expectile and Quantile Regression*. with contributions from P. H. C. Eilers, T. Kneib and G. Kauermann, R package version 0.36.

Stasinopoulos, M. and B. Rigby (2013). *gamlss: Generalized Additive Models for Location Scale and Shap*. R package version 4.2-6.

Tang, C. Y. and C. Leng (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika 98*(4), 1001–1006.

Taylor, J. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics 6*(2), 231–252.

Turlach, B. A. and A. Weingessel (2013). *quadprog: Functions to solve Quadratic Programming Problems*. R package version 1.5-5.

van Buuren, S. and A. Fredriks (2001). Worm plot: A simple diagnostic device for modeling growth reference curves. *Statistics in Medicine 20*, 1259–1277.

Verbeke, G. and G. Molenberghs (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The german socio-economic panel study (soep) - scope, evolution and enhancements. *Journal of Applied Social Science Studies 127*(1), 139–169.

Wand, M. (2013). *SemiPar: Semiparametic Regression*. R package version 1.0-4.

Wang, H. and X. Zhou (2010). Estimation of the retransformed conditional mean in health care cost studies. *Biometrika 97*(1), 147–158.

Winkler, R. L., G. M. Roodman, and R. R. Britney (1972). The determination of partial moments. *Management Science 19*, 290–296.

Wood, S. (2006). *Generalized Additive Models*. Boca Raton: Chapman and Hall.

Wood, S. (2013). *mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL*. R package version 1.7-24.

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

Yao, Q. and H. Tong (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of Nonparametric Statistics 6*(2-3), 273–292.

Yee, T. W. (2012). *VGAM: Vector Generalized Linear and Additive Models*. R package version 0.9-0.

Yuan, M. (2006). GACV for quantile smooting splines. *Computational Statistics and Data Analysis 50*(3), 813–829.

Ziegel, J. F. (2013). Coherence and elicitability. arXiv:1303.1690v2.

Eidesstattliche Versicherung

gemäß § 8 Abs. 2 Punkt 5 der Promotionsordnung

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 02.12.2014

(Linda Schulze Waltrup)