Partially Linear Censored Quantile Regression

Tereza Neocleous · Stephen Portnoy

Received: date / Accepted: date

Abstract Censored Regression Quantile (CRQ) methods provide a powerful and flexible approach to the analysis of censored survival data when standard linear models are felt to be appropriate. In many cases however, greater flexibility is desired to go beyond the usual multiple regression paradigm. One area of common interest is that of partially linear models: one (or more) of the explanatory covariates are assumed to act on the response through a non-linear function. Here the CRQ approach (Portnoy (2003)) is extended to this partially linear setting. Basic consistency results are presented. A simulation experiment and unemployment example justify the value of the partially linear approach over methods based on the Cox proportional hazards model and on methods not permitting nonlinearity.

Keywords quantile regression · partially linear models · B-splines · censored data · unemployment duration

1 Introduction

Consider the following data analysis problem: a large scale longitudinal survey is taken to study the durations of spells of unemployment of workers. Exits

---

Tereza Neocleous
Department of Statistics, University of Glasgow
15 University Gardens , Glasgow G12 8QW, UK
Tel.: +44-141-330-6117
Fax: +44-141-330-4814
E-mail: t.neocleous@stats.gla.ac.uk

Stephen Portnoy
Department of Statistics, University of Illinois
725 S. Wright St., Champaign IL 61801
Tel:+1-217-333-2167
Fax:+1-217-244-7190
E-mail: sportnoy@uiuc.edu

from unemployment to employment are marked and used to define observed periods of unemployment. Other exits are considered to generate censored values. A specific example is given in Section 5 with 2214 observed durations of which 55 % are censored. In addition to the unemployment durations, several covariates are observed: gender, marital status, place of residence, age, education (*etc.*). The basic models used for such data express the durations (or log-durations) as a linear model in the covariates (including their interactions).

As discussed below, censored regression quantile methods are especially appropriate since the relationships (that is, the parameters or the coefficients of linear regression terms) may be expected to vary with the size (conditional quantile) of the response or because of population heterogeneity. For example, the effect of nationality or gender may be quite different for people with short unemployment durations than for those with longer unemployment spells.

However, even at a fixed quantile, it seems highly unlikely that the effect of age is strictly linear (even if the data is transformed, say by logarithms). Thus, it is highly desirable to be able to allow the effect of age (and interactions with other covariates) to be modeled by somewhat nonlinear functions. An approach to providing such analyses is presented here.

We consider a regression quantile estimator for right censored survival data. Let $(\boldsymbol{X}, Y)$ be a random vector with $\boldsymbol{X} \in \mathbb{R}^{p'}$ and Y a real-valued variable. $\boldsymbol{X}$ could have discrete or continuous components, with at least one continuous component whose relationship with Y is nonlinear. For $\tau \in (0,1)$ the regression quantile $Q_{Y|\boldsymbol{X}}(\tau;\boldsymbol{x})$ of Y given $\boldsymbol{X}=\boldsymbol{x}$ satisfies

$$P(Y \le Q_{Y|\boldsymbol{X}}(\tau;\boldsymbol{x})|\boldsymbol{X}=\boldsymbol{x}) = \tau.$$

Assuming that n independent pairs $(Y_i, \boldsymbol{X}_i)$ are observed, and that the relationship between Y and $\boldsymbol{X}$ is linear, i.e.

$$Q_{Y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}(\tau), \tag{1}$$

the τth regression quantile coefficient, $\hat{\boldsymbol{\beta}}(\tau)$, and hence the regression quantile $\hat{Q}_{Y|\boldsymbol{X}}(\tau;\boldsymbol{x})$, can be obtained as the solution of

$$\min_{\boldsymbol{b}\in\mathbb{R}^{p'}} \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{b})$$

where $\rho_\tau(u) = u(\tau - I(u<0))$ (see Koenker (2005) for details). With survival times it is often the case that Y is not observed, and that instead one observes only the minimum of Y and a censoring variable C. Suppose that n independent triples $\{(\boldsymbol{X}_i, Z_i, \varDelta_i), i=1,\ldots,n\}$ are observed, with $Z_i = \min(Y_i, C_i)$ and $\varDelta_i = I(Y_i \le C_i)$. We are interested in estimating $Q_{Y|\boldsymbol{X}}(\tau;\boldsymbol{x})$ when Y and C are conditionally independent given $\boldsymbol{X}$, and when Y varies linearly with most components of $\boldsymbol{X}$ but nonlinearly with at least one component of $\boldsymbol{X}$.

Under the linear models paradigm a quantile regression approach is especially useful in survival analysis, as it interprets the covariate effect on survival times with flexibility not always achievable under the global assumptions like those of the Cox model. Koenker and Geling (2001) introduced a quantile regression approach to survival analysis by means of a transformation of the survival times. For instance, when the log-transformation is used, quantile regression corresponds to the accelerated failure time model, in which $\log Y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + u_i$ and the hazard rate is given by

$$h_i(y|\boldsymbol{x}_i) = h_0(y \exp(-\boldsymbol{x}_i^\intercal \boldsymbol{\beta})) \exp(-\boldsymbol{x}_i^\intercal \boldsymbol{\beta}).$$

Moreover, if the u_i are i.i.d. with extreme value distribution $F(u) = 1 - \exp(-\exp(u))$, this corresponds to the Cox proportional hazards model with Weibull baseline hazard, and the linear quantile regression model for the log-survival times agrees with the Cox model for accelerated failure time. Otherwise the Cox model specifies a parametric model for the survival distribution, while quantile regression permits rather general heterogeneity (subject to the use of linear models). The proportional hazards model is the most popular method for analyzing right-censored survival data, but in recent years there have been advances in quantile regression methods that offer an alternative to the Cox approach.

The earliest proposed estimator for censored quantile regression assumed fixed censoring (Powell (1986)). Subsequent research either assumed fixed censoring or independence between Y and C, *e.g.* Buchinsky and Hahn (1998), Honore et al (2002), and Chernozhukov and Hong (2002).

The independence assumption was relaxed in Portnoy (2003), where conditional independence of Y and C given $\boldsymbol{x}$ is assumed, and a "reweighting-to-the-right" (Efron (1967)) scheme is employed to compute the conditional quantiles. The Portnoy (2003) method is of particular interest, as it essentially extends the Kaplan-Meier estimator to the regression setting. A similar generalization of the Nelson-Aalen estimator was also recently proposed by Peng and Huang (2008). The models developed in the rest of this paper are based on the Portnoy estimator.

The Portnoy CRQ model assumes conditional independence between Y_i and C_i given $\boldsymbol{x}_i$. The approach is based on a recursive pivoting algorithm for random censoring, whose solution reduces to the Kaplan-Meier estimator in the one-sample case. The algorithm iteratively computes the entire conditional quantile function for $\tau \in (0, 1)$, stopping at a value of τ for which all observations remaining above the current conditional quantile function are censored. Note that this differs from the usual quantile regression methods that compute the conditional quantile at a fixed τ. If, for instance, the median is required, the pivoting algorithm of Portnoy (2003) will compute all quantiles up to the 50th in order to obtain the median.

In what follows, we present a modification of the pivoting algorithm with a generalization permitting nonlinear response to one (or more) covariates (as a "partially linear" model). Section 2 presents a grid algorithm as a computationally effective method for fitting such models based on generally available regression quantile programs. Section 3 examines the asymptotic properties of the partially linear CRQ estimator, extending earlier results for linear CRQ estimators given in Vanden Branden (2005), Neocleous et al (2006), and Portnoy and Lin (2008). Simulation experiments are statistically analyzed in Section 4 to evaluate the performance of the approach. A study of unemployment duration data is presented in Section 5 to show the value of the use of the partially linear censored regression model.

2 Grid algorithm for linear CRQ

A slightly modified version of the Portnoy (2003) CRQ pivoting algorithm, evaluating the linear regression quantiles of (1) on a grid of τ values is presented here. This algorithm iteratively computes the conditional quantiles from lowest to highest. Suppose that at the starting value t_1 of $\tau \in (0,1)$ there are no censored observations below the t_1th quantile, so that the quantile coefficient $\hat{\boldsymbol{\beta}}(t_1)$ is estimated using the usual quantile regression algorithm minimizing $\sum_{i=1}^{n} \rho_{t_1}(y_i - \boldsymbol{x}_i^\intercal \boldsymbol{b})$ with respect to $\boldsymbol{b}$. The corresponding quantile hyperplane $\boldsymbol{x}_i^\intercal \hat{\boldsymbol{\beta}}(t_1)$ will then have proportion t_1 of the data below it and $(1-t_1)$ above. We say that observations for which $Y_i \le \boldsymbol{x}_i^\intercal \hat{\boldsymbol{\beta}}(t_1)$ *are crossed* by the t_1th quantile. As the value of τ increases, censored observations may also get crossed. When the ith censored observation is crossed, the algorithm splits it to two parts according to a weighting scheme: a part that is observed at C_i and a part at infinity. If the ith censored point C_i is crossed for the first time at $\tau = \tau_i(\boldsymbol{\beta})$, it will receive weight

$$\hat{w}_i(\tau, \boldsymbol{\beta}) = (\tau - \tau_i(\boldsymbol{\beta}))/(1 - \tau_i(\boldsymbol{\beta})) \tag{2}$$

for all $\tau > \tau_i(\boldsymbol{\beta})$. This weight is updated every time τ increases. With weights for all crossed censored observations computed, weighted quantile regression is performed to obtain the regression coefficients at the current value of τ. More details on the weights of crossed observations and on the weighted quantile regression performed are given below.

Algorithm

1. Choose gridpoints $t_1, \ldots, t_M$ covering the set $\varepsilon \le \tau \le 1$. Starting with the gridpoint t_1 compute the initial quantile function $\hat{\boldsymbol{\beta}}(t_1)$ for $1 \le \tau \le t_1$ using the uncensored quantile regression algorithm applied to all the observations (both censored and uncensored). This assumes that the initial regression quantile, $\hat{\boldsymbol{\beta}}(t_1)$, determines a hyperplane that lies below all censored points, which is reasonable since censored observations below all data are non-informative and can be deleted without changing the estimation (as is the case in the Kaplan-Meier estimator).
2. Suppose that the quantiles $\hat{\boldsymbol{\beta}}(t_l), 1 \le l \le k$ have been computed and that weights $\hat{w}_i(\tau, \hat{\boldsymbol{\beta}})$ (see (2)) have been computed using these $\hat{\boldsymbol{\beta}}$ estimates.

Find $\hat{\boldsymbol{\beta}}(t_{k+1})$ by minimizing over $b \in \mathbb{R}^{p'}$ the objective function

$$\sum_{i=1}^{n} \Big\{ \Delta_i \rho_{t_{k+1}}(Y_i - \boldsymbol{x}_i^\intercal \boldsymbol{b}) + (1-\Delta_i)\big[\hat{w}_i(t_{k+1}, \boldsymbol{\beta})\rho_{t_{k+1}}(C_i - \boldsymbol{x}_i^\intercal \boldsymbol{b}) + (1-\hat{w}_i(t_{k+1}, \boldsymbol{\beta}))\rho_{t_{k+1}}(Y^* - \boldsymbol{x}_i^\intercal \boldsymbol{b})\big]\Big\}$$

where Y^* is a sufficiently large value so that $Y^* > \boldsymbol{x}_i^\intercal \boldsymbol{b}$ for all $\boldsymbol{x}_i^\intercal \boldsymbol{b}$ from the data. Y^* will be referred to as "point at infinity".

3. In the step from t_k to t_{k+1} some censored observations that were not previously crossed might get crossed. For those observations we have that $C_i > x_i^\intercal \hat{\boldsymbol{\beta}}(t_k)$ and $C_i \leq x_i^\intercal \hat{\boldsymbol{\beta}}(t_{k+1})$. They are then given weights $\hat{w}_i(\tau, \hat{\boldsymbol{\beta}}) = (\tau - \tau_i(\hat{\boldsymbol{\beta}}))/(1-\tau_i(\hat{\boldsymbol{\beta}}))$ with $\tau_i(\hat{\boldsymbol{\beta}}) = t_{k+1}$ with the rest of the weight going to the point at infinity, Y^*. In addition, updated weights are computed for the already crossed observations according to formula (2). With all the weights defined, another weighted quantile regression is performed as in step 2 above at $\tau = t_{k+2}$.
4. The algorithm stops either at the last grid point, t_M, or at some point t_e when only non-reweighted censored observations remain above the current solution, $\boldsymbol{x}_i^\intercal \hat{\boldsymbol{\beta}}(t_e)$.

The main advantage of using the grid modification of the pivoting algorithm is computational. For large sample sizes the pivoting algorithm computes solutions at a high number of τ-values. With the grid algorithm the number of τ-values at which the solution is obtained can be reduced, with substantial savings in computational time required for the iterative process. The grid algorithm is outlined above for a linear CRQ model. In what follows the algorithm is applied within the framework of partially linear models.

3 The partially linear estimator and its large sample properties

The partially linear CRQ model combines semiparametric estimation for censored data with quantile regression techniques, and uses B-splines for the estimation of the nonlinear term. Consider first the uncensored fully nonlinear model $y_i = g_\tau(x_i) + e_i$, where the e_i are independent random errors with τth quantile equal to zero. Following the notation in Schumaker (1981), let

$$\pi(s) = (B_1(s), B_2(s), \ldots, B_{k'_n+d+1}(s))^\intercal$$

be the set of B-spline basis functions with given knots $\Delta = \{z_i\}_0^{k'_n}$ with number of spline knots ${k_n}'$ and order of splines $d+1$. Then the estimated τth quantile function $\hat{g}_{n\tau}(s) = \pi(s)^\intercal \hat{\boldsymbol{\theta}}_n$, where $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{k'_n+d+1}$, is a solution of

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{k'_n+d+1}} \sum_i \rho(y_i - \pi(x_i)^\intercal \boldsymbol{\theta}).$$

Once the spline knots are selected and the spline bases computed, the problem is reduced to a linear quantile regression with $(k_n'+d+1)$ parameters. In the uncensored case it was shown, *e.g.* in He and Shi (1994, 1996) that if g_τ is smooth with bounded rth derivative, and ${k_n}'$ is of order $n^{1/(2r+1)}$, under some mild conditions the spline estimate $\hat{g}_{n\tau}(s)$ converges to $g_\tau(s)$ at the optimal nonparametric rate of $\mathcal{O}_p(n^{-2r/(2r+1)})$. In what follows we discuss the use of a B-spline estimator in a censored regression quantile setting.

Assuming the data $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$, $i = 1, \ldots, n$, come from a model with

$$Q_{Y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_{1i}^\mathsf{T}\theta_1(\tau) + g_\tau(\boldsymbol{x}_{2i}), \tag{3}$$

the estimated quantiles will be of the form

$$\hat{Q}_{Y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_{1i}^\mathsf{T}\hat{\theta}_1(\tau) + \pi(\boldsymbol{x}_{2i})^\mathsf{T}\hat{\theta}_2(\tau), \tag{4}$$

where g_τ is approximated by a linear combination of B-splines.

Let $\boldsymbol{\beta} = (\boldsymbol{\theta}_1^\mathsf{T}, \boldsymbol{\theta}_2^\mathsf{T})^\mathsf{T}$. Without loss of generality, we assume that the support of $g(s)$ is $s \in [0,1]$. Let $\pi(s) = (\pi_1(s), \pi_2(s), \ldots, \pi_{k_n'+d+1}(s))^\mathsf{T}$ be the B-spline basis of order d with $k_n^{'}$ knots. Let $k_n = k_n^{'} + d + 1$ and define $R_i(\tau) = \pi(\boldsymbol{x}_{2i})^\mathsf{T}\theta_2(\tau) - g_\tau(\boldsymbol{x}_{2i})$. Then at the kth step of the CRQ grid algorithm the estimated t_{k+1}th quantile is $\boldsymbol{x}_{1i}^\mathsf{T}\hat{\boldsymbol{\theta}}_1(t_{k+1}) + \pi(\boldsymbol{x}_{2i})^\mathsf{T}\hat{\boldsymbol{\theta}}_2(t_{k+1})$. This linearity in $\boldsymbol{\beta}$ allows current theoretical approaches to be generalized to the case of $\boldsymbol{\beta}$ of increasing dimension (at the same rate as k_n). For a grid of M τ-values the CRQ estimator is $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}(t_1)^\mathsf{T}, \hat{\boldsymbol{\beta}}(t_2)^\mathsf{T}, \ldots, \hat{\boldsymbol{\beta}}(t_M)^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{Mp}$, where $p = q + k_n$ and $q = \dim(\boldsymbol{\theta}_1)$. Asymptotic results for the linear CRQ model presented in Vanden Branden (2005), Neocleous et al (2006), and Portnoy and Lin (2008) are extended to the partially linear CRQ model as follows:

Theorem 1 *Let $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{Mp}$, be the censored regression quantile estimator for the model specified in (1) on a grid $\varepsilon \leqslant t_1 < t_2 < \ldots < t_M \leqslant 1-\varepsilon$. Let $\boldsymbol{\beta}^*$ be the true unknown censored regression quantile coefficient along the same grid, $t_{k+1} - t_k \equiv g_n = n^{-\kappa}$ and $k_n = \mathcal{O}(n^\gamma)$ where γ and κ satisfy one of (5), (6) and (7):*

$$0 < \kappa < 1/6, \qquad 0 < \gamma < \kappa \tag{5}$$

$$1/6 < \kappa < 1/4, \qquad 0 < \gamma < 1/4 \tag{6}$$

$$1/4 < \kappa < 1/3, \qquad 0 < \gamma < 1/2(1-3\kappa). \tag{7}$$

Under Assumptions (I), (F), (X) and (XX) given in the Appendix,

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|^2 = \mathcal{O}_p(n^{\kappa+\gamma-1}).$$

For the partially linear CRQ model with B-spline estimation of the nonlinear part, the following corollary holds.

Corollary 1 *Let $\hat{\beta} = (\hat{\theta}_1^\mathsf{T}, \hat{\theta}_2^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{Mp}$ be the censored regression quantile grid estimator of $\beta^* = (\theta_1^{*\mathsf{T}}, \theta_2^{*\mathsf{T}})$, where $\pi(x_2)^\mathsf{T}\theta_2^*$ estimates $g(x_2)$ in the model specified in (3). Under the assumptions and notation of Theorem 1, with the added condition*

(G) $g_\tau(s)$ *has bounded* r*th derivative for* $r \geq 3$ *for all* τ,

$$\|\hat{\theta}_1 - \theta_1^*\|^2 = \mathcal{O}_p(n^{\kappa+\gamma-1}).$$

Corollary 1 can be proved by combining B-spline approximation rates and Theorem 1. This result is most useful in applications where the effect of interest, *e.g.* treatment effect, is to be estimated in the presence of some additional nonlinear covariate.

4 Simulation study

To examine the finite sample performance of the partially linear CRQ estimator, we conducted a simulation experiment in which the censored response is linear in one covariate and non-linear in another covariate. Event times were generated for $i = 1, \ldots, n$ from the model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \frac{10 e_{1i}}{1 + \exp(6 - 0.5 x_{2i})}$$

and censoring times from the model (Configuration 1)

$$C_i = \beta_0 + \beta_1 x_{1i} + \frac{10 e_{2i}}{1 + \exp(5 - 0.5 x_{2i})}$$

for roughly 20% censoring, or (Configuration 2)

$$C_i = \beta_0 + \beta_1 x_{1i} + \frac{10 e_{2i}}{1 + \exp(4 - x_{2i})} - 0.2 x_{1i}^2$$

for roughly 40% censoring. Parameter values were $\beta_0 = 1$ and $\beta_1 = 3$, and the x_{1i} were generated as iid $U(0, 5)$, the x_{2i} as iid $U(0, 25)$, and e_{1i} and e_{2i} as iid $N(1, 0.01)$. The scatterplot in Figure 1 shows the censoring mechanism for Configuration 1 and sample size $n = 500$. Four different models were fitted to the data: one with linear term in x_2 and three with spline terms of order 2, 3 and 4 (piecewise linear, quadratic and cubic) in x_2. Knots at the quartiles of x_2 were used in the spline models for Configuration 1, while for Configuration 2 two additional sets of knots for x_2 were considered. In each case bootstrap confidence intervals were computed with $b = 500$ bootstrap replications.

Tables 1 and 2 report average bias, median absolute error, root mean square error, empirical coverage probability (95% nominal coverage) and mean confidence interval length for the slope of x_1 evaluated at $\tau = 0.50$ and 0.75 (similar results were obtained for $\tau = 0.25$) for Configuration 1. In all cases the partially linear model outperforms its linear equivalent. The difference between the three spline orders used is less clear, with some evidence that the quadratic spline works best. This is also supported by Figure 2, in which the quadratic spline term appears to give the best fit for the nonlinear term.

The effect of knot selection and placement is further investigated in the simulation study of Configuration 2, in which fitted spline models have knots at (a) the 33rd and 66th quantile of x_2, (b) the quartiles of x_2, and (c) the 20th, 40th, 60th and 80th quantiles of x_2. Tables 3 and 4 show the performance of various models fitted for Configuration 2. It is seen that again the spline models perform better than the linear model, while three knots are in general better than just two. The difference between three and four knots is less clear, as it appears that three knots are better for quadratic spline models, and four knots better for piecewise linear and cubic spline models.

Finally, Table 5 reports bias and root mean square error for the estimation of the nonlinear term in Configuration 2. The quadratic spline with three knots appears to be performing better than other spline models in terms of root mean square error. Differences in bias are less obvious.

TABLES 1-5 ABOUT HERE.

5 Application to unemployment duration

We illustrate the usefulness of the partially linear CRQ model with an application to administrative unemployment data from the German Socio-Economic Panel Survey, a longitudinal survey of private households in Germany covering topics such as income, employment, education and health. We focus on a subset of the data covering the period 1992-2004. The response variable of interest, Y, is the duration in months of the latest unemployment spell in the respondent's work history.

We restrict our attention to males with German nationality (as both nationality and gender were found to be significant in preliminary analyses) and we explore the effect of age and marital status on unemployment duration. Exits from unemployment to full- or part-time employment were considered observed while all other exits were considered as censored observations. Excluding observations with missing data, this gave a sample size of 2214 records with 55% censoring. Of these 2214 individuals, 42% were married. The median age for married respondents was 47.42 and for single 26.17.

The CRQ model

$$Q_{\log(Y)}(\tau \mid \boldsymbol{x}) = \beta_0(\tau) + \beta_1(\tau) \times \text{ married} + \boldsymbol{\theta}(\tau)^\top \pi(\text{age}) \tag{8}$$

was considered and quantiles up to the 60th were estimated. In particular, a quadratic spline term with knots at the quartiles of age was fitted. This provides a smooth 5-parameter fit to the age effect. All but one of the five coefficients were significant (at some τ-values), and so it is clear that the age effect requires more than a linear term.

Plots of $\hat{\beta}(\tau)$, the estimated quantile coefficients for the intercept and marital status, against τ are shown in Figure 3. The coefficients tend to be smaller in absolute value for short term unemployment (lower τ values) and larger for long term unemployment (higher τ values).

Marriage has a strong negative effect on unemployment duration, independent of age (the relevant interaction terms were not significant). The estimated median coefficient representing the difference in log-duration between a single and a married German male is -0.8244 (confidence interval of (-1.1649,-0.4838)), i.e. median unemployment duration for a married respondent is 0.4385 times that of a single respondent of the same age. The size of the marriage effect is similar in all but the lowest quantiles of unemployment duration.

Plots of the estimated median unemployment duration against age are shown in Figure 4 separately for single and married German males. Pointwise bootstrap confidence intervals are also shown. The age ranges plotted reflect the different age distributions for married and single groups. For married males over 50, censoring exceeds 80%, thus we restrict attention for the married group to the "reliable estimation" age range (31.42,50.00) corresponding to the 10th age percentile and the age with 81% censoring above it. For single males the age range plotted is (19.67,47.17) corresponding to the 10th and 90th age percentiles. In the singles age distribution, 80% of the observations over age 47.17 are censored.

From Figure 4, it is clear that the age effect on unemployment duration is quite nonlinear (at least for single men), with age being beneficial at very low ages (< 25) and rather detrimental (for both single and married men) at higher ages (as might be expected). The quantile analysis in Figure 5 presents perhaps a more surprising result. For quantiles below $\tau = 0.3$ (shorter unemployment durations), the effect is rather independent of age. This is not unexpected, as those who are readily re-employable do well at any age. However, for higher quantiles, the detrimental effect of age seems to increase rapidly for men in the range 30 - 50 years. The rather substantial increase in difficulty for older men who are not so readily re-employable would seem to call for some explanation (economic, psychological, or sociological).

Plots such as those in Figures 3-5 are useful in identifying departures from linearity. We advocate exploring the nonlinearity of each continuous covariate before attempting to fit linear coefficients as a way to detect patterns and improve the overall fit of the model. In addition, fitting a CRQ model is especially useful in highlighting population heterogeneity that is reflected in different structures for covariate effects for long and short durations. Such heterogeneity can not be identified in general with proportional hazards models. While proportional hazards models with time-varying coefficients have been proposed (see Gray (1992) or Tian et al (2005) for a more recent example), such models focus on dynamic structural change and do not provide estimates of regression effects on specific population quantiles. That is, while such models can identify secular trends, they can not identify structural variability for different population quantiles (as do CRQ models). It may be noted that such models are inherently nonparametric, even in the absence of partially linear covariate effects, with strictly slower convergence than the root-n asymptotics of standard regression quantile methods. On the other hand, time-varying covariates or coefficients can not be incorporated directly into CRQ models (nor into accelerated failure time models in general), and so CRQ models can not identify secular trends effectively.

6 Concluding remarks

In the preceding sections the use of a partially linear model for censored regression quantiles was proposed as a useful extension to standard linear regression techniques for survival data. The partially linear model was shown to be consistent and its use was illustrated by a data example and simulations. Quartile knots were used for the B-spline estimation of nonlinear terms and the quadratic spline gave satisfactory quantile estimates in the empirical example and simulations. Higher order spline terms did not show much improvement in estimation.

The censored regression quantile estimator is robust and flexible enough to highlight aspects of the data that the most common survival analysis techniques might overlook. Incorporating a nonlinear part adds even more flexibility to the model, allowing for more accurate estimation of parameters of interest, like quantile treatment effects. Censored regression quantiles and the semiparametric model proposed here are tools for capturing subtle aspects of the data and can be used in conjunction with other techniques for more comprehensive exploration of censored data.

The partially linear CRQ model can be extended to accommodate more than one nonlinear effects, as the basic theory extends directly to higher dimensions. However the curse of dimensionality could make application to two or more covariates quite problematic, in terms of slow convergence, complicated choice of a large number of knots, and interpretability.

As in every semiparametric model, the use of B-splines raises the question of knot selection. In this work the spline knots were chosen at fixed quantiles of the nonlinear variable. As long as the knot selection is not data-driven (*e.g.* equally spaced knots or quantile knots, perhaps depending on the sample size n), the asymptotic theory of B-splines applies directly (and consistency follows by Theorem 1 if the number of knots increases with n appropriately). Asymptotic results are not currently available if knot selection is data-driven. In practice fixing knots at specified quantiles of the x-variable is a simple and convenient solution for small to medium-sized datasets, and it is not likely that data-driven methods can offer much improvement here. However, in general it is also desirable to have a method for optimal knot selection and placement depending on the data. Such methods have been proposed by a number of authors. For instance, Koenker et al (1994) use a roughness penalty for quantile smoothing splines, and Doksum and Koo (2000) propose a method for stepwise knot addition and deletion using modified AIC and BIC for nonparametric quantile regression with regression splines. Further work along such lines would be useful for larger data sets.

Acknowledgments

This research was partially supported by National Science Foundation Grant DMS06-04229 and by the European Commission under the 6th Framework Programme's Research Infrastructures Action (Trans-national Access contract RITA 026040) hosted by IRISS-C/I at CEPS/INSTEAD, Differdange (Luxembourg). The data used in this publication were made available to us by the German Socio-Economic Panel Study (SOEP) at the German Institute for Economic Research (DIW), Berlin.

References

Buchinsky M, Hahn JY (1998) An alternative estimator for the censored quantile regression model. Econometrica 66:653–671

Chernozhukov V, Hong H (2002) Three-step censored quantile regression and extramarital affairs. J Am Stat Assoc 97:872–882

Doksum K, Koo J (2000) On spline estimators and prediction intervals in nonparametric regression. Comp Stat and Data Anal 35:67–82

Efron B (1967) The two-sample problem with censored data. In: Proceedings of the Fifth Berkeley Symposium, vol 4, pp 831–853

Gray RJ (1992) Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. J Am Stat Assoc 87:942–950

He X, Shao Q (2000) On parameters of increasing dimensions. J Multivariate Anal 73:120–135

He X, Shi P (1994) Convergence rate of B-spline estimators of nonparametric conditional quantile functions. Nonparametric Statistics 3:299–308

He X, Shi P (1996) Bivariate tensor-product B-splines in a partly linear model. J Multivariate Anal 58:162–181

Honore B, Khan S, Powell JL (2002) Quantile regression under random censoring. J Econometrics 109:67–105

Koenker R (2005) Quantile Regression. Cambridge University Press

Koenker R, Geling O (2001) Reappraising medfly longevity: a quantile regression survival analysis. J Am Stat Assoc 96:458–468

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

Neocleous T, Vanden Branden K, Portnoy S (2006) Correction to: "Censored Regression Quantiles". J Am Stat Assoc 101:860–861

Peng L, Huang Y (2008) Survival analysis with quantile regression models. J Am Stat Assoc 103:637–649

Portnoy S (2003) Censored regression quantiles. J Am Stat Assoc 98:1001–1012

Portnoy S, Lin G (2008) Asymptotics for censored regression quantiles. Submitted

Powell JL (1986) Censored regression quantiles. J Econometrics 32:143–155

Tian L, Zucker D, Wei LJ (2005) On the Cox model with time-varying regression coefficients. J Am Stat Assoc 100:172–183

Vanden Branden K (2005) Robust methods for high-dimensional data, and a theoretical study of depth-related estimators. PhD thesis, Katholieke Universiteit Leuven

Appendix: Proof of Theorem 1

The conditions for the main result (Theorem 1) are as follows:

(I) Y and C are conditionally independent given $\boldsymbol{x}$

(F) For $0 < \varepsilon < 1$, there exist constants a_j, b_j, c_j with $a_j > 0$ and $b_j < \infty$ for $j = 1, 2, 3$ such that

$$\begin{array}{ll} a_1 \leq f_{Y_i}(y) \leq b_1 & |f'_{Y_i}(y)| \leq c_1 \\ a_2 \leq \tilde{f}_{Y_i}(u) \leq b_2 & |\tilde{f}'_{Y_i}(u)| \leq c_2 \\ a_3 \leq \tilde{f}_{C_i}(v) \leq b_3 & |\tilde{f}'_{C_i}(v)| \leq c_3 \end{array}$$

uniformly for $\varepsilon \leq F_{Y_i}(y) \leq 1-\varepsilon$, $\varepsilon \leq \tilde{F}_{Y_i}(u) \leq 1-\varepsilon$ and $\varepsilon \leq \tilde{F}_{C_i}(v) \leq 1-\varepsilon$ and uniformly in $i = 1, \ldots, n$.

(X) $\max_{1 \leq i \leq n} ||\boldsymbol{x}_i|| = \mathcal{O}(p)$.

(XX) The matrix $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\intercal}$ is positive definite.

Theorem 1 makes use of the theory of He and Shao (2000) on the asymptotics of M-estimators when the parameter dimension increases with n. Briefly, this is outlined as follows. Let $\hat{\boldsymbol{\beta}}_n \in \mathbb{R}^m$ be the M-estimator for minimizing $\sum_{i=1}^{n} \zeta(\boldsymbol{z}_i, \boldsymbol{\beta})$ for some data set $\{\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n\}$ with $\boldsymbol{z}_i \in \mathbb{R}^{p+1}$ for $i = 1, 2, \ldots, n$; and for some objective kernel $\zeta(\boldsymbol{z}_i, \boldsymbol{\beta})$. If the objective function is convex in $\boldsymbol{\beta}$, and if $\zeta(\boldsymbol{z}, \boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\beta}$, except at finitely many points, with derivative $\Psi(\boldsymbol{z}, \boldsymbol{\beta})$, then Theorem 2.1 of He and Shao (2000) states that under certain conditions, $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2 = \mathcal{O}_p(m/n)$ where $\boldsymbol{\beta}^*$ is the solution to $\sum_{i=1}^{n} E_{\boldsymbol{\beta}} \Psi(\boldsymbol{z}_i, \boldsymbol{\beta}) = 0$. For the CRQ grid estimator the increasing dimension is $m = Mp$, where M is the number of grid points. Let $p = \mathcal{O}(n^\gamma)$ for some $\gamma > 0$. Equivalently, $p \leq cn^\gamma$ for some constant c. Define $\Psi_k(\boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{x}_i\{\Delta_i(I(Y_i < \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}(t_k)) + (1-\Delta_i)(w_i(\boldsymbol{\beta}, t_k)I(C_i < \boldsymbol{x}_i^{\intercal}\boldsymbol{\beta}(t_k)) - t_k)\}$,

$$\eta_i(\boldsymbol{\beta}', \boldsymbol{\beta}) = \Psi(\boldsymbol{x}_i, \boldsymbol{\beta}') - \Psi(\boldsymbol{x}_i, \boldsymbol{\beta}) - E(\Psi(\boldsymbol{x}_i, \boldsymbol{\beta}') - \Psi(\boldsymbol{x}_i, \boldsymbol{\beta}))$$

and $S_m = \{\alpha \in \mathbb{R}^m : \|\alpha\| = 1\}$. Then

$$\Psi(\boldsymbol{x}_i, \boldsymbol{\beta}) = (\Psi_1(\boldsymbol{x}_i, \boldsymbol{\beta})^{\intercal}, \Psi_2(\boldsymbol{x}_i, \boldsymbol{\beta})^{\intercal}, \ldots, \Psi_M(\boldsymbol{x}_i, \boldsymbol{\beta})^{\intercal})^{\intercal} \in \mathbb{R}^m.$$

The result also relies on the following two lemmas, which have been shown in the case of fixed p by Vanden Branden (2005). Here the result is extended to the case of p growing with n. Lemma 1 permits restricting the proof to monotone functions $\boldsymbol{x}^{\intercal}\boldsymbol{\beta}(\tau)$ on the grid. Lemma 2 shows that $\tau_i(\boldsymbol{\beta})$ and $\tau_i(\boldsymbol{\beta}^*)$ are close on the set of slopes $\boldsymbol{\beta}$ considered.

Lemma 1 *For every $B>0$, $\exists\, n_0$ such that for $n \geqslant n_0$ the set*

$$\left\{\boldsymbol{\beta} \in \mathbb{R}^m : \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leqslant B\left(\frac{m}{n}\right)^{1/2}\right\}$$

is contained in the set of all monotonic functions on the grid $\varepsilon \leqslant t_1 < t_2 < \ldots \leqslant t_M \leqslant 1-\varepsilon$ for some $\varepsilon > 0$. Here $t_k - t_{k-1} = g_n = n^{-\kappa}$, $p \leq cn^{\gamma}$ for some $c > 0$, and $m \leqslant p/g_n$, with $\gamma \leq \frac{1}{2} - \frac{3\kappa}{2}$, $\kappa > 0$.

Lemma 2 *Let $\tau_i(\boldsymbol{\beta})$ be the gridpoint at which $\boldsymbol{\beta}$ crosses C_i, and let $\tau_i(\boldsymbol{\beta}^*)$ be the unknown gridpoint at which the true regression quantile $\boldsymbol{\beta}^*$ crosses the same observation. It then holds that*

$$|\tau_i(\boldsymbol{\beta}) - \tau_i(\boldsymbol{\beta}^*)| = \mathcal{O}(T(n,m))$$

on the set $\{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leqslant B(m/n)^{1/2}\}$ with

$$T(n,m) = \max(Bc^{1/2}p^{1/2}(m/n)^{1/2}, 2g_n) = \max(Bcn^{\kappa+\gamma-1/2}, 2n^{-\kappa}).$$

Proofs of Lemmas 1 and 2 are straightforward generalizations of those in Vanden Branden (2005).

Proof (Proof of Theorem 1) It is sufficient to verify the following conditions of He and Shao (2000).

(C0) $\|\sum_{i=1}^n \Psi(\boldsymbol{x}_i, \hat{\boldsymbol{\beta}}_n)\| = o_p(n^{1/2})$.

(C1) There exists a C and $r \in (0,2]$ such that

$$\max_{i \leqslant n} E_{\boldsymbol{\beta}} \sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\beta}\| \leqslant d} \|\eta_i(\boldsymbol{\theta}, \boldsymbol{\beta})\|^2 \leqslant n^C d^r$$

for $0 < d \leqslant 1$.

(C2) $\|\sum_{i=1}^n \Psi(\boldsymbol{x}_i, \boldsymbol{\beta}^*)\| = \mathcal{O}_p(nm)^{1/2}$ or $\sum_{i=1}^n E\|\Psi(\boldsymbol{x}_i, \boldsymbol{\beta}^*)\|^2 = \mathcal{O}(nm)$.

(C3) There exists a sequence of $(m \times m)$ matrices D_n with $\liminf_{n\to\infty} \lambda_{\min}(D_n) > 0$ (where $\lambda_{\min}$ denotes the minimum eigenvalue) such that for any $B > 0$ and uniformly in $\boldsymbol{\alpha} \in S_m$

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\| \leqslant B(\frac{m}{n})^{1/2}} |\boldsymbol{\alpha}^{\mathsf{T}} \sum_{i=1}^n E_{\boldsymbol{\beta}^*}(\Psi(\boldsymbol{x}_i, \boldsymbol{\beta}) - \Psi(\boldsymbol{x}_i, \boldsymbol{\beta}^*)) - n\boldsymbol{\alpha}^{\mathsf{T}} D_n(\boldsymbol{\beta} - \boldsymbol{\beta}^*)| = o(n^{1/2}).$$

(C4) There exists a sequence $A(n,m) = o(n/\log n)$ for which

$$\sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\| \leqslant B(\frac{m}{n})^{1/2}} \sum_{i=1}^n E_{\boldsymbol{\beta}} |\boldsymbol{\alpha}^{\mathsf{T}} \eta_i(\boldsymbol{\beta}, \boldsymbol{\beta}^*)|^2 = \mathcal{O}(A(n,m))$$

for any $\boldsymbol{\alpha} \in S_m$, and $B > 0$.

(C5) $\sup_{\boldsymbol{\alpha} \in S_m} \sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\beta}^*\| \leqslant B(\frac{m}{n})^{1/2}} \sum_{i=1}^n (\boldsymbol{\alpha}^{\mathsf{T}} \eta_i(\boldsymbol{\beta}, \boldsymbol{\beta}^*))^2 = \mathcal{O}_p(A(n,m))$ for any $B > 0$.

(C0) follows from the gradient conditions by noting that

$$||\Psi(\hat{\boldsymbol{\beta}})||^2 = \mathcal{O}_P(M \max_{1\le k\le M} ||\Psi_k(\hat{\boldsymbol{\beta}}(t_k))||^2)$$

and

$$||\Psi_k(\hat{\boldsymbol{\beta}})|| = \mathcal{O}_P(\sqrt{p\log n}\max||\boldsymbol{x}_i||).$$

Thus

$$||\Psi(\hat{\boldsymbol{\beta}})|| = \mathcal{O}_P(p\sqrt{M\log n}) = \mathcal{O}_P(n^{\kappa/2+\gamma}(\log n)^{1/2}).$$

This is $o_p(n^{1/2})$, provided that $\kappa/2+\gamma<1/2$.

For (C1), we note that had the $\boldsymbol{x}_i$ been bounded by a constant, then $E_{\boldsymbol{\beta}}\,||\eta_{i,k}(\boldsymbol{\theta},\boldsymbol{\beta})||^2$ would have been bounded by a constant also. Since $\max||\boldsymbol{x}_i||^2 = \mathcal{O}(p)$, then $E_{\boldsymbol{\beta}}\,||\eta_{i,k}(\boldsymbol{\theta},\boldsymbol{\beta})||^2 = \mathcal{O}(p)$ and $E_{\boldsymbol{\beta}}\,||\eta_i(\boldsymbol{\theta},\boldsymbol{\beta})||^2 = \mathcal{O}(Mp)$, where $Mp\le cn^{\kappa+\gamma}$. Therefore one can take n large enough such that $C>\kappa+\gamma$ is satisfied with $0<d\le 1$. For (C2), we note that $E\,||\Psi_k(\boldsymbol{\beta}^*)||^2 = \mathcal{O}(\max||\boldsymbol{x}_i||^2)$ and

$$\sum_{i=1}^n\sum_{k=1}^M E\,||\Psi_k(\boldsymbol{\beta}^*)||^2 = \mathcal{O}(Mnp) = \mathcal{O}(mn).$$

(C3) and (C4) are the hardest conditions to prove. As shown in Vanden Branden (2005), for $\boldsymbol{\alpha}\in S_m$,

$$\boldsymbol{\alpha}^\intercal E\left[\Psi(\boldsymbol{\beta})-\Psi(\boldsymbol{\beta}^*)\right] = n\boldsymbol{\alpha}^\intercal D_n(\boldsymbol{\beta}-\boldsymbol{\beta}^*) \tag{9}$$

$$+\sum_{i=1}^n\sum_{k=1}^M \boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i\left\{\tilde{f}'_{Y_i}(u)(\boldsymbol{x}_i^\intercal(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))^2\right\} \tag{10}$$

$$+\sum_{i=1}^n\sum_{k=1}^M \boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i\left\{\sum_{l=1}^k d_{kl}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^\intercal(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2\right\} \tag{11}$$

where

$$d_{kl} = \begin{cases} -w_1 & l=1 \\ w_{k-1} & l=k \\ -(w_l-w_{l-1}) & otherwise \end{cases}$$

$$d_{kli} = \begin{cases} d_{kk}\tilde{f}_{C_i}(\boldsymbol{x}_i^\intercal\boldsymbol{\beta}^*(t_k)) + \tilde{f}_{Y_i}(\boldsymbol{x}_i^\intercal\boldsymbol{\beta}^*(t_k)) & l=k \\ d_{kl}\tilde{f}_{C_i}(\boldsymbol{x}_i^\intercal\boldsymbol{\beta}^*(t_l)) & otherwise \end{cases}$$

and

$$nD_n = \begin{pmatrix} \sum_{i=1}^{n} d_{11i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \mathbf{0}_{p,p} & \dots & \dots & \dots & \dots & \mathbf{0}_{p,p} \\ \sum_{i=1}^{n} d_{21i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \sum_{i=1}^{n} d_{22i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \dots & \dots & \dots & \dots & \mathbf{0}_{p,p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^{n} d_{k1i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \sum_{i=1}^{n} d_{k2i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \dots & \sum_{i=1}^{n} d_{kki}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \mathbf{0}_{p,p} & \dots & \mathbf{0}_{p,p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^{n} d_{M1i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \sum_{i=1}^{n} d_{M2i}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} & \dots & \dots & \dots & \dots & \sum_{i=1}^{n} d_{MMi}\boldsymbol{x}_i\boldsymbol{x}_i^{\intercal} \end{pmatrix}. \tag{12}$$

Thus for (C3) to hold we require

$$|\sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\left\{\tilde{f}'_{Y_i}(u)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))^2+\sum_{l=1}^{k}d_{kl}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2\right\}| \\ = o(n^{1/2}) \tag{13}$$

or, as noted in Remark 2.3 of He and Shao (2000),

$$|\sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\left\{\tilde{f}'_{Y_i}(u)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))^2+\sum_{l=1}^{k}d_{kl}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2\right\}| \\ = o((mn)^{1/2}). \tag{14}$$

For(10) we have

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\tilde{f}'_{Y_i}(u)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))^2 \\ &\leq \sum_{i=1}^{n}\sum_{k=1}^{M}|\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i|\sum_{k=1}^{M}(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))^2 \\ &\leq \sum_{i=1}^{n}||\boldsymbol{x}_i||\left(\sum_{k=1}^{M}||\boldsymbol{\alpha}_i||^2\right)^{1/2}||\boldsymbol{x}_i||^2\sum_{k=1}^{M}||\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)||^2 \\ &= \mathcal{O}(\frac{m}{n}\sum_{i=1}^{n}||\boldsymbol{x}_i||^3) = \mathcal{O}(p^{3/2}m) = \mathcal{O}(n^{5\gamma/2+\kappa}) \end{aligned}$$

and for (11)

$$\sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\sum_{l=1}^{k}d_{kl}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2$$
$$\leq \sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i d_{k1}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2 \tag{15}$$
$$+\sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\sum_{l=2}^{k}d_{kl}\tilde{f}'_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2 \tag{16}$$

Noting that $d_{kl} = \mathcal{O}(1)$ for $l = 1$ and $\mathcal{O}(M)$ otherwise, we obtain

$$\begin{aligned}(15) &\leq \sum_{i=1}^{n}\left(\sum_{k=1}^{M}||\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i||^2\right)^{1/2}\left(\sum_{k=1}^{M}d_{k1}^2\right)^{1/2}||\boldsymbol{x}_i||^2\,||\boldsymbol{\beta}(t_1)-\boldsymbol{\beta}^*(t_1)||^2\\
&\leq \sum_{i=1}^{n}||\boldsymbol{x}_i||^3\,M^{1/2}\,||\boldsymbol{\beta}(t_1)-\boldsymbol{\beta}^*(t_1)||^2 = O(p^{3/2}M^{1/2}\frac{m}{n})\\
&= O(n^{5\gamma/2+3\kappa/2})\end{aligned}$$

and

$$\begin{aligned}(16) &\leq \sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\left(\sum_{l=2}^{k}d_{kl}^2\right)^{1/2}\sum_{l=2}^{k}(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))^2\\
&\leq \sum_{i=1}^{n}\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\left(\frac{k-1}{M^2}\right)^{1/2}||\boldsymbol{x}_i||^2\sum_{l=2}^{k}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l))^2\\
&= \mathcal{O}(M^{1/2}\frac{m}{n}\sum_{i=1}^{n}||\boldsymbol{x}_i||^3) = \mathcal{O}(p^{3/2}M^{1/2}\frac{m}{n})\\
&= \mathcal{O}(n^{5\gamma/2+3\kappa/2}).\end{aligned}$$

With γ and κ satisfying $2\gamma + \kappa < 1/2$, the error from (C3) can be made $o((mn)^{1/2}) = o(n^{\gamma/2+\kappa/2+1/2})$.

(C4) needs to hold with $A(n, m) = o(n/\log n)$. The term $\boldsymbol{\alpha}^{\intercal}\eta_i(\boldsymbol{\beta}, \boldsymbol{\beta}^*)$ is defined as

$$\boldsymbol{\alpha}^{\intercal}\eta_i(\boldsymbol{\beta},\boldsymbol{\beta}^*) = \sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}(\Psi_k(\boldsymbol{x}_i,\boldsymbol{\beta})-\Psi_k(\boldsymbol{x}_i,\boldsymbol{\beta}^*)-E(\Psi_k(\boldsymbol{x}_i,\boldsymbol{\beta})-\Psi_k(\boldsymbol{x}_i,\boldsymbol{\beta}^*))). \tag{17}$$

A Taylor series expansion for the expectation part of the expression gives

$$\sum_{k=1}^{M}\boldsymbol{\alpha}_k^{\intercal}\boldsymbol{x}_i\left[\tilde{f}_{Y_i}(u)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}^*(t_k)))+\sum_{l=1}^{k}d_{kl}\tilde{f}_{C_i}(v)(\boldsymbol{x}_i^{\intercal}(\boldsymbol{\beta}(t_l)-\boldsymbol{\beta}^*(t_l)))\right]$$

for some u and v. Similarly as for (C3) the first part of this term is bounded by $\mathcal{O}((m/n)^{1/2}p) = \mathcal{O}(n^{\kappa/2+3\gamma/2-1/2})$ and the second part is bounded by $\mathcal{O}(p(Mm/n)^{1/2}) = \mathcal{O}(n^{3\gamma/2+\kappa/2-1/2})$. Therefore

$$\boldsymbol{\alpha}^\intercal \eta_i(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = \sum_{k=1}^{M} \boldsymbol{\alpha}_k^\intercal (\Psi_k(\boldsymbol{x}_i, \boldsymbol{\beta}) - \Psi_k(\boldsymbol{x}_i, \boldsymbol{\beta}^*)) + \mathcal{O}(n^{3\gamma/2+\kappa/2-1/2}).$$

This error term squared and multiplied by n is $\mathcal{O}(n^{3\gamma+\kappa})$ which can be made $o(n/\log n)$ if $3\gamma + \kappa < 1$ so that it satisfies the requirement for (C4). For the term in $\sum_{k=1}^{M} \boldsymbol{\alpha}_k^\intercal (\Psi_k(\boldsymbol{x}_i, \boldsymbol{\beta}) - \Psi_k(\boldsymbol{x}_i, \boldsymbol{\beta}^*))$ we introduce an indicator, $I_{a_k,b_k}(Y)$, with $I_{a_k,b_k}(Y) = \pm 1$ if Y lies in between $\boldsymbol{x}_i^\intercal a(t_k)$ and $\boldsymbol{x}_i^\intercal b(t_k)$, and 0 otherwise. Then

$$\begin{aligned}
&\sum_{k=1}^{M} \boldsymbol{\alpha}_k^\intercal (\Psi_{ki}(\boldsymbol{\beta}) - \Psi_{ki}(\boldsymbol{\beta}^*)) = \sum_{k=1}^{M} \boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i \operatorname{sign}\big(()\boldsymbol{x}_i^\intercal (\boldsymbol{\beta}(t_k) - \boldsymbol{\beta}^*(t_k))\times \\
&[I(Y_i \leqslant C_i) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(Y_i) + I(Y_i > C_i) w(\boldsymbol{\beta}^*, t_k) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(C_i)] \\
&\qquad + \sum_{k=1}^{M} \boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i I(Y_i > C_i) I(C_i \leqslant \boldsymbol{x}_i^\intercal \boldsymbol{\beta}(t_k))(w_i(\boldsymbol{\beta}, t_k) - w_i(\boldsymbol{\beta}^*, t_k)).
\end{aligned}$$

The last term can be bounded using Lemma 2. For some constant D

$$|w_i(\boldsymbol{\beta}, t_k) - w_i(\boldsymbol{\beta}^*, t_k)| = \left| \frac{(t_k - 1)(\tau_i(\boldsymbol{\beta}) - \tau_i(\boldsymbol{\beta}^*))}{(1 - \tau_i(\boldsymbol{\beta}))(1 - \tau_i(\boldsymbol{\beta}^*))} \right| \leqslant DT(n, m)$$

where $T(n, m)$ is as defined in Lemma 2. Therefore the last term can be bounded by

$$\begin{aligned}
\mathcal{O}(M^{1/2} p^{1/2} T(n, m)) &= \max\big(\mathcal{O}(M^{1/2} p (m/n)^{1/2}), \mathcal{O}(p^{1/2}/M^{1/2})\big) \\
&= \max(\mathcal{O}(n^{3\gamma/2+\kappa-1/2}), \mathcal{O}(n^{\gamma/2-\kappa/2})).
\end{aligned}$$

Combining these results gives

$$\begin{aligned}
|\boldsymbol{\alpha}^\intercal \eta_i(\boldsymbol{\beta}, \boldsymbol{\beta}^*)| \leqslant &\sum_{k=1}^{M} \{|\boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i| [|I(Y_i \leqslant C_i) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(Y_i)| + |I(Y_i > C_i) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(C_i)|\}] \\
&+ \max\Big(\mathcal{O}(n^{3\gamma/2+\kappa-1/2}), \mathcal{O}(n^{\gamma/2-\kappa/2})\Big).
\end{aligned}$$

This error term squared and multiplied by n will be $o(n/\log n)$ if $3\gamma + 2\kappa < 1$ and $\gamma - \kappa < 0$.

Finally for the term

$$\left(\sum_{k=1}^{M} \{|\boldsymbol{\alpha}_k^\intercal \boldsymbol{x}_i| [|I(Y_i \leqslant C_i) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(Y_i)| + |I(Y_i > C_i) I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(C_i)|\}] \right)^2,$$

a bound is required on the number of observations for which $I_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_k^*}(Y_i)$ and $I_{\boldsymbol{\beta}_l, \boldsymbol{\beta}_l^*}(Y_i)$ with $l \neq k$ are both non-zero. By Lemma 2, this number is bounded by $D^*T(n,m)M$ for some constant D^*. A bound of $\mathcal{O}(p(m/n)^{1/2}) = \mathcal{O}(n^{3\gamma/2+\kappa/2-1/2})$ is thus obtained for the main part of the square. The cross term contributes

$$\mathcal{O}(p(m/n)^{1/2}T(n,m)M) = \max\left(\mathcal{O}(n^{5\gamma/2+5\kappa/2-1}), \mathcal{O}(n^{3\gamma/2+\kappa/2-1/2})\right).$$

The contribution of both terms can once again be made $o(n/\log n)$ if $5\gamma/2 + 5\kappa/2 < 1$ and $3\gamma/2 + \kappa/2 < 1/2$.

The constraints on κ and γ yield equations (5), (6) and (7).
All that is left is to verify that (C5) holds for these values.

According to Lemma 2.2 of He and Shao (2000), (C5) holds with the same $A(n,m)$ as in (C4), provided that $c_{n,m}^2 m \log n = \mathcal{O}(A(n,m))$, where $c_{n,m}$ is a sequence satisfying $\sup_{\boldsymbol{\beta},\boldsymbol{x}} \|\Psi(\boldsymbol{x},\boldsymbol{\beta})\| \leqslant c_{n,m}$. Here $c_{n,m} = D^{**}M^{1/2}p^{1/2}$ for some constant D^{**}. Recalling that $p = \mathcal{O}(n^\gamma)$, it follows that $c_{n,m}^2 m \log n = \mathcal{O}(A(n,m))$, which concludes the proof of Theorem 1.

Remark. The results obtained in Theorem 1 are not optimal. For example, one possible choice for γ and κ is $\gamma = 1/7$ and $\kappa = 1/5$ which would give a rate of order $n^{-23/35}$. In addition, if condition (C4) holds with $A(n,m) = o(\frac{n}{m \log n})$, Theorem 2.2 of He and Shao (2000) gives asymptotic normality of the estimator, but requires tighter bounds than those obtained in Vanden Branden (2005), Neocleous et al (2006) and in Theorem 1. That is not to say that asymptotic normality is not possible. In fact, empirical results show that as the sample size n increases, the distribution of the CRQ-estimated $\hat{\boldsymbol{\beta}}$ appears to approach a normal distribution.

Table 1 Comparison of performance for $\beta_1(0.50)$ in the simulation model with approximate 20% censoring (Configuration 1). Knots at the quartiles of x_2 were used for the spline terms. The average bias, median absolute error, root mean square error, empirical coverage probability (95% nominal coverage) and mean confidence interval length are shown.

$\tau = 0.50$	Bias	MAE	RMSE	ECP	EML
n=200					
lin	−0.00188	0.07646	0.11086	0.940	0.45406
pcs	−0.00012	0.00413	0.01115	0.996	0.04806
quad	0.00033	0.00436	0.00997	0.980	0.03552
cub	0.00024	0.00831	0.01420	0.968	0.05564
n=500					
lin	0.00262	0.05208	0.07554	0.936	0.28953
pcs	0.00003	0.00216	0.00419	0.990	0.01669
quad	−0.00019	0.00228	0.00452	0.950	0.01692
cub	−0.00003	0.00573	0.00843	0.960	0.03405
n=1000					
lin	−0.00198	0.03420	0.04850	0.952	0.20286
pcs	0.00001	0.00124	0.00228	0.982	0.00934
quad	−0.00011	0.00158	0.00291	0.950	0.01088
cub	−0.00005	0.00420	0.00609	0.954	0.02488

Table 2 Comparison of performance for $\beta_1(0.75)$ in the simulation model with approximate 20% censoring (Configuration 1). Knots at the quartiles of x_2 were used for the spline terms. The average bias, median absolute error, root mean square error, empirical coverage probability (95% nominal coverage) and mean confidence interval length are shown.

$\tau = 0.75$	Bias	MAE	RMSE	ECP	EML
n=200					
lin	−0.00167	0.06349	0.10313	0.928	0.40821
pcs	0.00081	0.00784	0.01576	0.969	0.05667
quad	−0.00004	0.00332	0.00787	0.994	0.03060
cub	0.00033	0.00637	0.01171	0.969	0.05071
n=500					
lin	0.00349	0.04290	0.06439	0.940	0.25481
pcs	−0.00001	0.00436	0.00771	0.949	0.02945
quad	−0.00014	0.00169	0.00352	0.978	0.01349
cub	−0.00028	0.00411	0.00707	0.966	0.02916
n=1000					
lin	−0.00432	0.03272	0.04355	0.954	0.17951
pcs	−0.00017	0.00353	0.00508	0.946	0.02003
quad	−0.00005	0.00124	0.00209	0.964	0.00815
cub	−0.00002	0.00302	0.00480	0.968	0.01980

Table 3 Comparison of performance for $\beta_1(0.50)$ in the simulation model with $n = 500$ and approximate 40% censoring (Configuration 2). Knots at (a) the 33rd and 66th quantiles, (b) the quartiles and (c) the 20th, 40th, 60th and 80th quantiles of x_2 were used for the spline terms. The average bias, median absolute error, root mean square error, empirical coverage probability (95% nominal coverage) and mean confidence interval length are shown.

$\tau = 0.50$	Bias	MAE	RMSE	ECP	EML
Linear term in x_2					
	−0.1074	0.1069	0.1256	0.6640	0.2835
Piecewise linear spline					
(a)	−0.0166	0.0173	0.0233	0.7980	0.0645
(b)	0.0108	0.0109	0.0212	0.9457	0.0641
(c)	0.0056	0.0081	0.0144	0.9618	0.0526
Quadratic spline					
(a)	0.0276	0.0288	0.0348	0.7560	0.0917
(b)	0.0010	0.0032	0.0055	0.9739	0.0219
(c)	0.0030	0.0047	0.0081	0.9379	0.0279
Cubic spline					
(a)	0.0018	0.0038	0.0060	0.9700	0.0242
(b)	0.0061	0.0080	0.0110	0.9280	0.0379
(c)	0.0008	0.0026	0.0040	0.9699	0.0172

Table 4 Comparison of performance for $\beta_1(0.75)$ in the simulation model with $n = 500$ and approximate 40% censoring (Configuration 2). Knots at (a) the 33rd and 66th quantiles, (b) the quartiles and (c) the 20th, 40th, 60th and 80th quantiles of x_2 were used for the spline terms. The average bias, median absolute error, root mean square error, empirical coverage probability (95% nominal coverage) and mean confidence interval length are shown.

$\tau = 0.75$	Bias	MAE	RMSE	ECP	EML
Linear term in x_2					
	−0.2084	0.2116	0.2239	0.3260	0.3246
Piecewise linear spline					
(a)	−0.0247	0.0253	0.0330	0.7818	0.0918
(b)	−0.0033	0.0091	0.0135	0.9277	0.0491
(c)	0.0023	0.0052	0.0093	0.9351	0.0361
Quadratic spline					
(a)	0.0111	0.0104	0.0159	0.8741	0.0500
(b)	0.0011	0.0033	0.0050	0.9834	0.0210
(c)	0.0021	0.0048	0.0077	0.9436	0.0289
Cubic spline					
(a)	0.0013	0.0040	0.0063	0.9529	0.0246
(b)	0.0035	0.0052	0.0081	0.9306	0.0306
(c)	0.0011	0.0029	0.0044	0.9741	0.0176

Table 5 Comparison of performance for $Q(\tau \mid x)$ in the simulation model with $n = 500$ and approximate 40% censoring (Configuration 2). Knots at (a) the 33rd and 66th quantiles, (b) the quartiles and (c) the 20th, 40th, 60th and 80th quantiles of x_2 were used for the spline terms. The root mean square error and average bias are shown for the 50th and 75th conditional quantiles evaluated at the mean of x_1.

	$\tau = 0.50$		$\tau = 0.75$	
	RMSE	Bias	RMSE	Bias
Linear term in x_2				
	1.3968	0.3233	1.3275	−0.0271
Piecewise linear spline				
(a)	0.4967	0.2426	0.4875	−0.1078
(b)	0.6699	0.3322	0.6281	−0.0182
(c)	0.4775	0.2711	0.4842	−0.0793
Quadratic spline				
(a)	0.8891	0.4601	0.7975	0.1096
(b)	0.4731	0.2755	0.4721	−0.0750
(c)	0.5090	0.2804	0.4973	−0.0700
Cubic spline				
(a)	0.4784	0.2682	0.4766	−0.0822
(b)	0.6032	0.3072	0.5558	−0.0432
(c)	0.4956	0.2680	0.4991	−0.0824

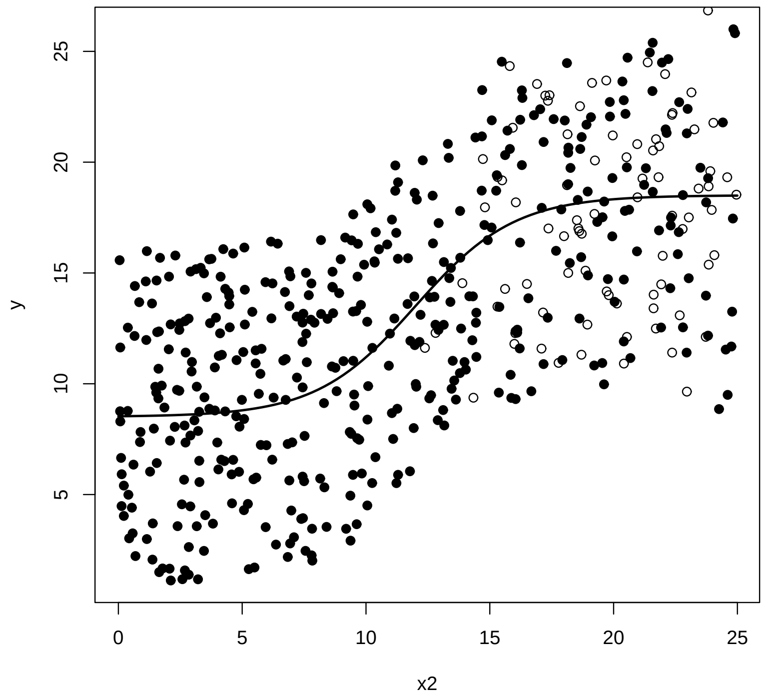

Fig. 1 Scatterplot of Configuration 1 used in the simulation experiment. Censored points are shown as open circles, uncensored points as filled circles. The conditional median line evaluated at the mean of x_1 is also shown.

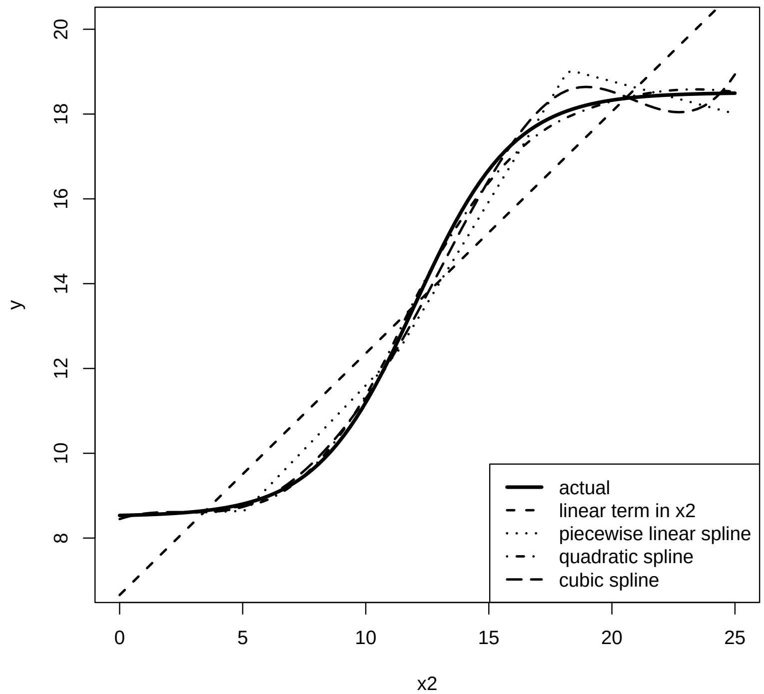

Fig. 2 Various model fits for the nonlinear term in the simulation experiment (Configuration 1). Shown here are the actual median (solid line) and model-estimated conditional median lines (dashed or dotted) evaluated at the mean of x_1.

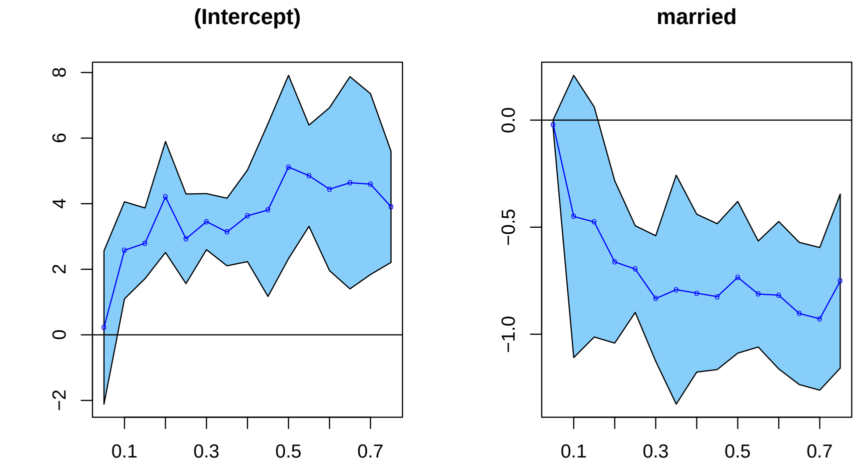

Fig. 3 Estimated linear coefficients $\hat{\beta}_0(\tau)$ and $\hat{\beta}_1(\tau)$ in model (8) with 95% bootstrap pointwise confidence intervals plotted against τ for $0 < \tau \leq 0.75$.

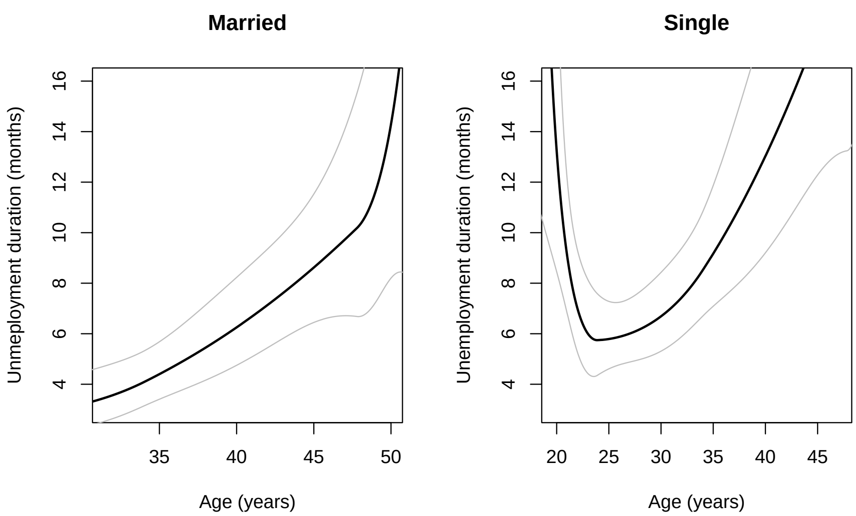

Fig. 4 Estimated median unemployment duration against age for German males. The black line shows the median, grey lines show 95% pointwise confidence limits.

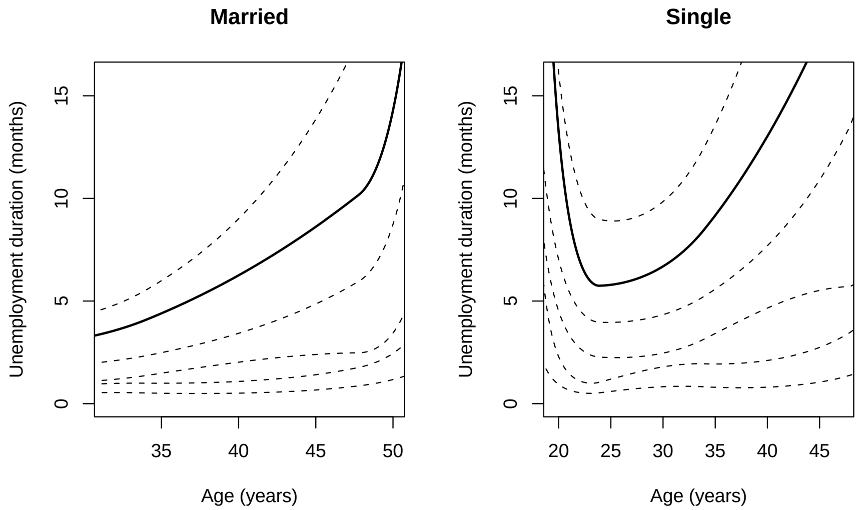

Fig. 5 Estimated deciles of unemployment duration against age for German males. The solid line shows the median, dashed lines show the other deciles from 1st to 6th.