

# Bias-corrected quantile regression estimation of censored regression models

P. Čížek<sup>1</sup> · S. Sadikoglu<sup>2</sup>

Received: 15 October 2014 / Revised: 4 March 2016 / Published online: 29 March 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** In this paper, an extension of the indirect inference methodology to semi-parametric estimation is explored in the context of censored regression. Motivated by weak small-sample performance of the censored regression quantile estimator proposed by Powell (J Econom 32:143–155, 1986a), two- and three-step estimation methods were introduced for estimation of the censored regression model under conditional quantile restriction. While those stepwise estimators have been proven to be consistent and asymptotically normal, their finite sample performance greatly depends on the specification of an initial estimator that selects the subsample to be used in subsequent steps. In this paper, an alternative semiparametric estimator is introduced that does not involve a selection procedure in the first step. The proposed estimator is based on the indirect inference principle and is shown to be consistent and asymptotically normal under appropriate regularity conditions. Its performance is demonstrated and compared to existing methods by means of Monte Carlo simulations.

**Keywords** Asymptotic normality · Censored regression · Indirect inference · Quantile regression

**Mathematics Subject Classification** 62F10 · 62F12 · 62N01 · 62N02

---

✉ P. Čížek  
p.cizek@uvt.nl  
S. Sadikoglu  
s.sadikoglu@uvt.nl

<sup>1</sup> CentER, Department of Econometrics & OR, Tilburg University, Tilburg, The Netherlands

<sup>2</sup> Department of Econometrics & OR, Tilburg University, Tilburg, The Netherlands

## 1 Introduction

The censored regression model has been studied and extensively used in a wide range of applied economics literature. To estimate the parameters of censored regression models, the maximum likelihood estimator (MLE) is usually used under the assumption that the error terms have distribution functions with a known parametric form (e.g., that the error terms are normally distributed). Contrary to the least squares estimator in linear regression model, the MLE is not robust to departures from the parametric assumptions about the error term distribution. If the employed assumptions do not hold, the MLE is in general inconsistent (cf. [Arabmazar and Schmidt 1982](#); [Brännäs and Laitila 1989](#)). The same applies also to Heckman's two-step estimator (see [Jonsson 2012](#)).

To relax the strong assumptions of MLE, several semiparametric estimators have been introduced in the econometric literature. Relying on very weak identification assumptions, [Powell \(1984, 1986a\)](#) proposed the censored least absolute deviation (CLAD) and censored regression quantile (CRQ) estimators by imposing the restriction that the conditional quantile of the error term is zero. These consistent and asymptotically normal estimators were applied in many contexts (e.g., [Fahr 2004](#); [Melenberg and van Soest 1996](#)), and furthermore, have been extended in many directions, which include random censoring ([Honore et al. 2002](#); [Portnoy 2003](#)) as well as panel-data models ([Honore 1992](#); [Campbell and Honore 1993](#)). In practice, the CRQ estimator is very appealing due to its robustness to misspecification of the error-term distribution and of the form of heteroskedasticity.

On the other hand, CRQ is difficult to compute exactly since its objective function is non-differentiable, non-convex, and more importantly, exhibits multiple local minima. The results using algorithms searching local minima (see [Fitzenberger 1997b](#), for an overview) depend on the choice of the starting points, while the algorithms searching the global minimum (e.g., [Fitzenberger and Winker 2007](#)) can become relatively demanding in terms of computation time, especially if many regressors are involved. In Monte Carlo experiments, [Fitzenberger \(1997a\)](#) demonstrated a more severe drawback of the CRQ estimator in small samples than the mean-biasedness and inefficiency documented already by [Paarsch \(1984\)](#) and [Moon \(1989\)](#): the CRQ estimator exhibits a heavy-tailed asymmetric distribution in small samples. These observations made also by [Fitzenberger and Winker \(2007\)](#), for instance, are supported by larger differences between expected and median values of CRQ in some simple settings, and as argued by [Khan and Powell \(2001\)](#), can be attributed to the fact that the CRQ parameter estimates play two roles: while they are estimates of the regression parameters, they also determine which observations have positive conditional quantiles and can thus enter the minimization of the CRQ objective function. (Note that these unfavorable finite-sample properties are shared to some extent also by some alternatives to CLAD such as the symmetrically censored least squares of [Powell \(1986b\)](#). Other alternative estimator such as those by [Horowitz \(1986\)](#) and [Honore and Powell \(1994\)](#) do not exhibit such heavy-tailed finite-sample distributions, but require the error terms and the explanatory variables being independent, which is a rather strong assumption.)

Since [Khan and Powell \(2001\)](#) highlighted the inherent property of CLAD and CRQ that causes their poor small-sample performance—the joint identification of

observations entering the objective function and of the quantile regression line, several stepwise estimation procedures have been proposed, for example, by [Khan and Powell \(2001\)](#) and [Chernozhukov and Hong \(2002\)](#). These methods select first a subset of observations to identify the quantile regression line and then apply the quantile regression (QR) on the selected observations. The first step can be achieved, for example, by a nonparametric selection procedure as in [Buchinsky and Hahn \(1998\)](#) and [Khan and Powell \(2001\)](#). Although asymptotically nearly equivalent to an ‘oracle’ QR estimator, the selection procedure in the first step works at a cost in finite samples. Alternatively, [Chernozhukov and Hong \(2002\)](#) developed a three-step estimation method that involves a parametric first step to circumvent the “curse-of-dimensionality” problem posed by nonparametric selection procedures. Its performance in small samples does not however improve upon the two-step estimators in simple regression models.

The finite-sample performance of the stepwise estimators does not seem substantially better than CLAD in studies of [Khan and Powell \(2001\)](#), [Chernozhukov and Hong \(2002\)](#), and most recently [Tang et al. \(2011\)](#). This might be due to relatively low precision of the initial nonparametric fit, for instance, but also due to a comparison with CLAD estimates based on favorable local minima rather than the global minimum. To improve upon the stepwise estimators, we introduce an alternative semiparametric estimator for the censored regression model under conditional quantile restriction. Contrary to the existing methods, we apply the linear regression QR estimator to all data (rather than to a preselected subsample) and then correct its bias caused by censoring. For the bias correction, indirect inference (II), which was suggested by [Gouriéroux et al. \(1993\)](#), is used. The indirect inference methodology is a simulation-based technique that is essentially used for estimation of the parameters of correctly specified but intractable models, but it can be employed as a bias correction method too (e.g., [Gouriéroux et al. 2000](#); [Gouriéroux et al. 2010](#)).

Implementing the standard II approach requires knowledge of the error-term distribution at least up to a parametric form. To exploit only the conditional quantile restriction, we propose a new II methodology based on simulating the values of the error terms from a semiparametrically estimated distribution. The proposed II estimator is based on the standard linear QR (which allows for linear, quadratic, and polynomial functions of regressors) for two reasons. First, linear QR has desirable properties such as convexity of the objective function and a reasonably small variance in small samples. Second and more importantly, the properties of linear QR are known even under model misspecification (see [Angrist et al. 2006](#)) and can be used to construct a nonparametrically estimated error distribution for the II simulations and subsequent bias correction. Hence, the proposed bias-corrected QR procedure can be shown to be consistent and asymptotically normal. Its benefits in small samples are demonstrated by means of Monte Carlo simulations.

The remainder of the paper is organized as follows. Section 2 presents a review of relevant estimation methods of censored regression model and a brief overview of the indirect inference methodology. In Sect. 3, the proposed bias-corrected QR estimator is described in details. The asymptotic properties of the indirect estimator are discussed in Sect. 4 and the results of Monte Carlo experiments are presented in Sect. 5. Proofs are given in the appendices.

## 2 Estimation of censored regression model and indirect inference

The censored regression model and some relevant estimators are introduced in Sect. 2.1 and the indirect inference concept is described in Sect. 2.2.

### 2.1 Censored regression model

Let us define the censored regression model. First, data are supposed to be a random sample of size  $n \in N$  originating from a latent linear regression model

$$y_i^* = x_i^T \beta^0 + \varepsilon_i, \quad (1)$$

where  $y_i^* \in R$  is the latent dependent variable,  $x_i \in R^k$  is the vector of explanatory variables,  $\beta^0$  represents the  $k$ -dimensional parameter vector, and  $\varepsilon_i$  is the unobserved error term with its conditional  $\tau$ -quantile,  $\tau \in (0, 1)$ , being zero:  $q_\tau(\varepsilon_i | x_i) = 0$ . The observed responses  $y_i$  equal to  $y_i^*$  censored from below at some cut-off point  $cp_i$ :

$$y_i = \max \left\{ cp_i, x_i^T \beta^0 + \varepsilon_i \right\}. \quad (2)$$

We consider here only the case of fixed censoring with a known cut-off point  $cp_i \equiv cp$ , and without loss of generality,  $cp = 0$ . An extension to random censoring is possible by the procedure of [Honore et al. \(2002\)](#).

The CRQ estimator is an extension of the classical linear QR to the censored regression model under a conditional quantile restriction. Since the conditional quantile function of  $y_i$  in (2) is simply  $\max\{0, x_i^T \beta^0\}$ , [Powell \(1986a\)](#) proposed the CRQ estimator  $\hat{\beta}_n^{CRQ}$ :

$$\hat{\beta}_n^{CRQ} = \arg \min_{\beta \in B} \sum_{i=1}^n \rho_\tau \left( y_i - \max \{0, x_i^T \beta\} \right), \quad (3)$$

where  $B$  is a compact parameter space,  $\rho_\tau(z) = \{\tau - I(z \leq 0)\}z$  with  $\tau \in (0, 1)$ , and  $I(\cdot)$  denotes the indicator function. Note that CRQ can be interpreted as applying the linear QR estimator to the observations  $x_i$  with  $x_i^T \beta^0 > 0$ : only observations with  $x_i^T \beta^0 > 0$  carry information to identify and estimate the conditional quantiles of  $y_i^*$  given  $x_i$ ,  $x_i^T \beta^0 > 0$ , whereas observations with  $x_i^T \beta^0 \leq 0$  carry information that the conditional quantile of  $y_i^*$  is negative and the conditional quantile of  $y_i$  equals zero given  $x_i$ ,  $x_i^T \beta^0 \leq 0$ , but they do not identify the conditional quantile of  $y_i^*$  at  $x_i$  and their contributions to the objective function (3) are independent of  $\beta$  in a neighborhood of  $\beta^0$ . This leads then to a heavy-tailed small-sample distribution of CRQ.

To eliminate this property, [Khan and Powell \(2001\)](#) proposed a two-step estimation method. In the first step, the observations with  $x_i^T \beta^0 > 0$  are determined by an initial semiparametric or nonparametric estimation, and in the second step, the standard QR estimation is conducted on the selected observations. Nevertheless, the finite sample results of [Khan and Powell \(2001\)](#) do not seem to generate a substantial advantage

with respect to the CRQ estimator in terms of mean or median squared errors, possibly due to an imprecise selection of observations in the first step; alternatively, using a local rather than a global optimization algorithm for CRQ could have played a role.

Later, [Chernozhukov and Hong \(2002\)](#) proposed a semiparametric three-step estimator of the censored regression model under conditional quantile restriction. The initial subset of observations with  $x_i^T \beta^0 > 0$  is selected by a parametric binary-choice model (e.g., logit) and QR is used in the subsequent steps to obtain not only consistent estimates, but also a more precise selection of the observations with  $x_i^T \beta^0 > 0$ . Their finite-sample results are however not substantially better than those of the two-step procedures: while having a smaller mean bias in small samples, the three-step estimates often exhibit a larger mean squared errors (cf. [Tang et al. 2011](#), Sect. 5).

## 2.2 Parametric indirect inference

Our strategy for estimating the censored regression model will differ from the existing ones in that QR will be applied to all observations and its bias due to censoring will be corrected by means of the indirect inference (II). In this section, we therefore describe a general principle of (parametric) II introduced by [Gouriéroux et al. \(1993\)](#) and discuss how II can be applied as a bias correction method following [Gouriéroux et al. \(2000\)](#).

Consider a general model, for example, (1)–(2):

$$y_i = h(x_i, \varepsilon_i; \beta), \quad (4)$$

where  $y_i$  represents the response variable,  $x_i \in R^k$  is the vector of explanatory variables with a distribution function  $G^0(\cdot)$ ,  $\beta^0 \in B \subset R^k$  is the parameter vector, and  $\varepsilon_i$  is the unobserved error term with a known conditional distribution function  $F^0(\cdot|x_i)$  (a generalization to a nonparametrically estimated distribution function will follow in Sect. 3).

To implement II, an instrumental criterion, which is a function of the observations  $\{y_i, x_i\}_{i=1}^n$  and of an auxiliary parameter vector  $\theta \in \Theta \subset R^q$ ,  $q \geq k$ , has to be defined (e.g., linear QR applied to censored data). This criterion is minimized to estimate the auxiliary parameter vector:

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} Q_n(\{y_i\}_{i=1}^n, \theta)$$

(please note that the dependence on the explanatory variables  $\{x_i\}_{i=1}^n$  is kept implicit as we do not consider simulating values of  $x_i$ , but work conditionally on observed  $\{x_i\}_{i=1}^n$ ; see [Gouriéroux et al. 1993](#), for details).

The data-generating process is then fully determined by  $F^0$  and  $\beta^0$  and the instrumental criterion  $Q_n$  is assumed to converge asymptotically to a non-stochastic limit that has a unique minimum  $\theta^0$ :

$$\theta^0 = \arg \min_{\theta \in \Theta} Q_\infty(F^0, \beta^0, \theta).$$

Evaluating it at any  $F(\cdot|x_i)$  and  $\beta$  leads to the definition of the binding function  $b(F, \beta)$ :

$$b(F, \beta) = \arg \min_{\theta \in \Theta} Q_{\infty}(F, \beta, \theta), \quad (5)$$

which implies that  $\theta^0 = b(F^0, \beta^0)$ .

Under some regularity assumptions (see Assumptions A1–A4 in [Gouriéroux et al. 1993](#)),  $\hat{\theta}_n$  is a consistent estimator of  $\theta^0$ . Provided that  $b(F^0, \beta)$  is known and one-to-one, a consistent estimate  $\hat{\beta}_n$  of  $\beta^0$  would be defined as  $\hat{\beta}_n = b^{-1}(F^0, \hat{\theta}_n)$  ( $F^0$  is traditionally assumed to be fully known; auxiliary parameters of the error distribution have to be a part of the parameter vector  $\theta$ ). Since the binding function is often difficult to compute, [Gouriéroux et al. \(1993\)](#) defined a simulation-based procedure to estimate the parameter  $\beta^0$ .

Let  $\{\tilde{\varepsilon}^1, \dots, \tilde{\varepsilon}^S\}$  be  $S$  sets of error terms, where  $\tilde{\varepsilon}^s = \{\tilde{\varepsilon}_i^s\}_{i=1}^n$ ,  $s = 1, \dots, S$ , are simulated from  $F^0(\cdot|x_i)$ ,  $\tilde{\varepsilon}_i^s|x_i \sim F^0(\cdot|x_i)$ , assuming this distribution is fully known. Then for any given  $\beta$ , one can generate  $S$  sets of simulated paths  $\{\tilde{y}^1(\beta), \dots, \tilde{y}^S(\beta)\}$  using model (4), where  $\tilde{y}^s(\beta) = \{\tilde{y}_i^s(\beta)\}_{i=1}^n$  and  $\tilde{y}_i^s(\beta) = h(x_i, \tilde{\varepsilon}_i^s; \beta)$  conditional on  $x_i$  for  $s = 1, \dots, S$ . From these simulated samples,  $S$  auxiliary estimates can be computed:

$$\tilde{\theta}_n^s(\beta) = \arg \min_{\theta \in \Theta} Q_n(\{\tilde{y}_i^s(\beta)\}_{i=1}^n, \theta). \quad (6)$$

Under appropriate conditions,  $\tilde{\theta}_n^s(\beta)$  tends asymptotically to  $b(F^0, \beta)$ , which allows to define the indirect inference estimator in the following way:

$$\hat{\beta}_n^{II} = \arg \min_{\beta \in B} \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta) \right]^T \Omega \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta) \right], \quad (7)$$

where  $\Omega$  is a positive definite weighting matrix. This estimator can be shown to be consistent and asymptotically normal (see Proposition 1 and 3 in [Gouriéroux et al. 1993](#)). As in GMM estimation, the choice of  $\Omega$  does not affect the asymptotic distribution of the estimator if  $\dim(\beta) = \dim(\theta)$  and its choice will thus be asymptotically irrelevant.

To argue that II can be used as a bias correction technique, note that  $\beta$  can represent the parameter value in the original model (4) (e.g., censored regression (1)–(2)) and  $\theta$  the parameter value of the auxiliary biased criterion (e.g., linear QR applied to censored data). The binding function  $b(F^0, \beta)$  then maps the parameter values  $\beta$  to biased estimates  $\theta$  and its inverse  $\hat{\beta}_n^{II} = b^{-1}(F^0, \hat{\theta}_n)$  maps the biased estimates back to the parameters in the original model; see [Gouriéroux et al. \(2000\)](#) for details.

### 3 Semiparametric indirect inference for censored regression

In this subsection, we introduce the semiparametric indirect estimation procedure to estimate the parameter vector of the censored regression model under conditional

quantile restriction. As the linear quantile regression is used as an instrumental criterion and the distribution of  $\varepsilon_i$  is unknown, a crucial ingredient of the procedure is the behavior of QR under misspecification (this is true even if quadratic or polynomial regression functions are used due to the kink of the true quantile function). Angrist et al. (2006) characterize the QR vector under misspecification as a minimizer of a weighted mean-squared approximation to the true conditional quantile function, assuming the almost-sure existence of the conditional density of the dependent variable, or equivalently, of the error term. As this result does not directly apply to the censored regression model, we modify their result to accommodate the fact that the error distribution is not continuous. In particular, we characterize the QR estimates under misspecification in terms of the distribution function of the error term rather than its density as Angrist et al. (2006) did.

Let us first introduce necessary notation. The conditional quantile function of the dependent variable  $y_i$  is  $\max\{0, x_i^T \beta^0\}$ . For any quantile index  $\tau \in (0, 1)$ , the QR vector is defined by:

$$\theta^0 = \arg \min_{\theta \in \Theta} E \left[ \rho_\tau(y_i - x_i^T \theta) \right], \quad (8)$$

where  $\rho_\tau(z) = \{\tau - I(z \leq 0)\}z$ . Further, let  $\Delta(x_i, \beta^0, \theta)$  denote the QR specification error,  $\Delta(x_i, \beta^0, \theta) = x_i^T \theta - \max\{0, x_i^T \beta^0\}$ , and the observed residual  $u_i$  be defined as  $u_i = y_i - \max\{0, x_i^T \beta^0\}$ . Finally, let  $F_u(u|x_i)$  and  $F_y(y|x_i)$  be the conditional distribution functions of  $u_i$  and  $y_i$ , respectively, and let  $f_u(u|x_i)$  and  $f_y(y|x_i)$  denote the corresponding conditional densities whenever they exist. For continuously distributed latent error  $\varepsilon_i$  in model (1),  $f_u(u|x_i)$  and  $f_y(y|x_i)$  exist and will be used only for  $u_i > -\max\{0, x_i^T \beta^0\}$  and  $y_i > 0$ , respectively.

**Theorem 1** Suppose that  $E(y_i)$  and  $E\|x_i\|^2$  are finite,  $\theta^0$  uniquely solves (8), and  $P\{\Delta(x_i, \beta^0, \theta^0) = 0\} = 0$ . Then,  $\bar{\theta} = \theta^0$  uniquely solves the equation

$$\bar{\theta} = \arg \min_{\theta \in \Theta} E \left[ w(x_i, \beta^0, \bar{\theta}) \cdot \Delta^2(x_i, \beta^0, \theta) \right], \quad (9)$$

where

$$w(x_i, \beta^0, \bar{\theta}) = \begin{cases} \frac{F_u\{\Delta(x_i, \beta^0, \bar{\theta})|x_i\} - \tau}{2\Delta(x_i, \beta^0, \bar{\theta})} & \text{if } \Delta(x_i, \beta^0, \bar{\theta}) \neq 0 \\ w_0(x_i) & \text{if } \Delta(x_i, \beta^0, \bar{\theta}) = 0 \end{cases}$$

for any bounded function  $w_0(x_i): R^k \rightarrow R_0^+$ .

*Proof* See Appendix 1. □

Theorem 1 states that the linear QR vector depends on the weighting function  $w(x_i, \beta^0, \theta^0)$ , which in turn is a function of the distribution function  $F_u(\cdot|x_i)$ . Thus, for any other distribution function  $\tilde{F}_u(\cdot)$  such that  $\tilde{F}_u\{\Delta(x_i, \beta^0, \theta^0)|x_i\} =$

$F_u\{\Delta(x_i, \beta^0, \theta^0)|x_i\}$ , the weighting function remains unchanged and the linear QR yields the same vector  $\theta^0$ .

Next, we consider Theorem 1 in the censored regression model with errors  $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$ . To facilitate semiparametric estimation that does not require complete estimation of  $F_\varepsilon(\cdot|x_i)$ , we will construct another distribution  $\tilde{F}_\varepsilon(\cdot|x_i)$  that is always from the same parametric family of distributions and that results in the same QR fits as  $F_\varepsilon(\cdot|x_i)$  by Theorem 1.

First, note that  $u_i = \max\{\varepsilon_i, -x_i^T \beta^0\}$  for  $x_i^T \beta^0 > 0$  and  $u_i = y_i$  for  $x_i^T \beta^0 \leq 0$ . Considering now different errors  $\tilde{\varepsilon}_i \sim \tilde{F}_\varepsilon(\cdot|x_i)$ , one can set  $\tilde{u}_i = \max\{\tilde{\varepsilon}_i, -x_i^T \beta^0\}$  for  $x_i^T \beta^0 > 0$  and  $\tilde{u}_i = \tilde{y}_i = \max\{0, x_i^T \beta^0 + \tilde{\varepsilon}_i\}$  for  $x_i^T \beta^0 < 0$ . As the censoring points of  $u_i$  and  $\tilde{u}_i$  are identical (conditionally on  $x_i$ ), we essentially have to match the two latent continuous distributions of  $\varepsilon_i$  and  $\tilde{\varepsilon}_i$ . We will show that the distribution  $\tilde{\varepsilon}_i$  can be a normal one:  $\tilde{\varepsilon}_i \sim N(\mu_\tau \sigma(x_i; \beta^0), \sigma(x_i; \beta^0))$ , where  $\mu_\tau$  is the  $(1 - \tau)$ th conditional quantile of the standard normal distribution  $N(0, 1)$  and  $\sigma^2(x_i; \beta^0)$  denotes the conditional variance. Specifically, we find  $\sigma(x_i; \beta^0)$  such that  $\tilde{F}_u\{\Delta(x_i, \beta^0, \theta^0)|x_i\} = F_u\{\Delta(x_i, \beta^0, \theta^0)|x_i\}$  for any finite value of  $x_i$ . First note that  $F_u\{\Delta(x_i, \beta^0, \theta^0)|x_i\} = F_y(x_i^T \theta^0|x_i)$  (and analogously for  $\tilde{u}_i$  and  $\tilde{y}_i$ ):  $\sigma(x_i; \beta^0)$  has to be therefore chosen so that

$$F_y(x_i^T \theta^0|x_i) = \tilde{F}_{\tilde{y}}(x_i^T \theta^0|x_i). \quad (10)$$

The definition of  $\sigma(x_i; \beta^0)$  is irrelevant if  $x_i^T \theta^0 < 0$  as then  $F_y(x_i^T \theta^0|x_i) = \tilde{F}_{\tilde{y}}(x_i^T \theta^0|x_i) = 0$ . Ignoring the case of  $F_y(x_i^T \theta^0|x_i) = \tau$ , which will be dealt with later, (10) for  $x_i^T \theta^0 \geq 0$  means

$$F_y(x_i^T \theta^0|x_i) = \Phi\left(\frac{x_i^T \theta^0 - [x_i^T \beta^0 + \mu_\tau \sigma(x_i; \beta^0)]}{\sigma(x_i; \beta^0)}\right) = \Phi_\tau\left(\frac{x_i^T \theta^0 - x_i^T \beta^0}{\sigma(x_i; \beta^0)}\right) \quad (11)$$

and

$$\sigma(x_i; \beta^0) = \frac{x_i^T \theta^0 - x_i^T \beta^0}{\Phi_\tau^{-1}(F_y(x_i^T \theta^0|x_i))}, \quad (12)$$

where  $\Phi$  and  $\Phi_\tau$  are the distribution functions of  $N(0, 1)$  and  $N(\mu_\tau, 1)$ , respectively (note that (12) leads to  $\sigma(x_i, \beta^0) = 0$  for  $x_i^T \theta^0 < 0$ ). Therefore, having  $\tilde{\varepsilon}_i \sim N(\mu_\tau \cdot \sigma(x_i; \beta^0), \sigma(x_i; \beta^0))$  with  $\sigma(x_i, \beta^0)$  defined in (12) yields the same linear QR vector as the real data generated under  $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$  and the biases of the linear QR estimates in the censored regression model (1)–(2) both with the original data distribution  $\varepsilon_i \sim F_\varepsilon(\cdot|x_i)$  and with the data generated from  $\tilde{\varepsilon}_i \sim N(\mu_\tau \cdot \sigma(x_i; \beta^0), \sigma(x_i; \beta^0))$  will be equal.

For the case of  $F_y(x_i^T \theta^0|x_i) = \tau$ , i.e.,  $x_i^T(\theta^0 - \beta^0) \rightarrow 0$  for  $x_i^T \beta^0 > 0$ , we consider the limit of (12) and define  $\sigma(x_i; \beta^0)$  as



$$\begin{aligned}\sigma(x_i; \beta^0) &= \frac{\phi_\tau \{ \Phi_\tau^{-1} [F_y(x_i^T \beta^0 | x_i)] \}}{f_y(x_i^T \beta^0 | x_i)} \\ &= \frac{\phi_\tau \{ \Phi_\tau^{-1} [F_y(x_i^T \theta^0 | x_i)] \}}{f_y(x_i^T \theta^0 | x_i)} \quad \text{if } F_y(x_i^T \theta^0 | x_i) = \tau.\end{aligned}\quad (13)$$

If the bias correction of linear QR is to be performed by II, we can simulate the set of error terms  $\{\tilde{\varepsilon}^1, \dots, \tilde{\varepsilon}^S\}$  from  $N(\mu_\tau \cdot \sigma(x_i; \beta), \sigma(x_i; \beta))$  instead of the original data distribution and calibrate over  $\beta \in B$  (provided that  $\sigma(x_i, \beta)$  is known). However,  $\beta$  is not identified in this case because Eq. (11) holds for any value  $\beta$  substituted for  $\beta^0$  if definition (12) is used at that  $\beta$ . To achieve identification,  $\beta^0$  in (12) has to be replaced by an initial estimate or the denominator in (12) also has to be a function of  $\beta$  instead of being equal to its true value at  $\beta^0$ . We consider the latter strategy to achieve good performance even in very small samples. It is well known that the identification of the parameter vector in (1)–(2) under conditional quantile restriction relies on the observations with positive values of the index  $x_i^T \beta^0 > 0$  (Powell 1984, 1986a) since  $F_y(x_i^T \beta^0 | x_i) = \tau$  only if  $x_i^T \beta^0 > 0$ . We also exploit this fact and we define  $\tilde{\sigma}(x_i; \beta)$  for  $x_i^T \theta^0 \geq 0$  as

$$\tilde{\sigma}(x_i; \beta) = \begin{cases} \frac{x_i^T \theta^0 - x_i^T \beta}{\Phi_\tau^{-1}(F_{y_i}(x_i^T \theta^0 | x_i))} & \text{if } x_i^T \beta \leq 0, \\ \frac{x_i^T \theta^0 - x_i^T \beta}{\Phi_\tau^{-1}(\min\{\max\{F_y(x_i^T \theta^0 | x_i) - F_y(x_i^T \beta | x_i) + \tau, 0\}, 1\})} & \text{if } x_i^T \beta > 0; \end{cases} \quad (14)$$

( $\tilde{\sigma}(x_i; \beta) = 0$  for  $x_i^T \theta^0 < 0$ ).<sup>1</sup> Since  $\tilde{\sigma}(x_i; \beta) = \sigma(x_i; \beta)$  only if  $\beta = \beta^0$ , (11) using  $\tilde{\sigma}(x_i; \beta)$  will hold only at  $\beta \equiv \beta^0$  and the parameter vector  $\beta$  can be identified (see Lemma 1 for details). Further, as (14) becomes indeterminate if  $F_y(x_i^T \theta^0 | x_i) = \tau$  or  $F_y(x_i^T \theta^0 | x_i) = F_y(x_i^T \beta | x_i)$ , we replace the definition of  $\tilde{\sigma}(x_i; \beta)$  in such cases by (13) so that  $\tilde{\sigma}(x_i; \beta)$  is continuous in  $\beta$ :

$$\begin{aligned}\tilde{\sigma}(x_i; \beta) &= \frac{\phi_\tau \{ \Phi_\tau^{-1} [F_y(x_i^T \theta^0 | x_i)] \}}{f_y(x_i^T \theta^0 | x_i)} \quad \text{if } F_y(x_i^T \theta^0 | x_i) = \tau \\ &\quad \text{or } F_y(x_i^T \theta^0 | x_i) = F_y(x_i^T \beta | x_i),\end{aligned}\quad (15)$$

where  $\phi_\tau(\cdot)$  is the density function of  $\Phi_\tau(\cdot)$  (note that the limit is the same for both cases of (14)).

With the definition (14) and (15) of  $\tilde{\sigma}(x_i; \beta)$ , which assumes knowledge of the true conditional distribution  $F_y(x_i^T \theta^0 | x_i)$  at  $x_i^T \theta^0$ , we can define the infeasible indirect inference (III) estimator  $\hat{\beta}_n^{III}$  by

<sup>1</sup> For  $x_i^T \beta \leq 0$ ,  $\tilde{\sigma}(x_i; \beta)$  might take a negative value. However, note that the data-generating process for simulated data as well as the identification of  $\beta$  are invariant to the sign of  $\tilde{\sigma}(x_i; \beta)$ .

$$\widehat{\beta}_n^{III} = \arg \min_{\beta \in B} \left[ \widehat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \widetilde{\theta}_n^s(\beta) \right]^T \Omega \left[ \widehat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \widetilde{\theta}_n^s(\beta) \right], \quad (16)$$

where  $\widetilde{\theta}_n^s(\beta) = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho_\tau(\widetilde{y}_i^s(\beta) - x_i^T \theta)$  using simulated data  $\widetilde{y}_i^s(\beta) = \max\{0, x_i^T \beta + \widetilde{\varepsilon}_i^s\}$ ,  $\widetilde{\varepsilon}_i^s \sim N(\mu_\tau \cdot \widetilde{\sigma}(x_i; \beta^0), \widetilde{\sigma}(x_i; \beta^0))$  for  $s = 1, \dots, S$ , and  $\Omega$  is a positive definite weighting matrix. For the sake of brevity, these distributions  $N(\mu_\tau \cdot \widetilde{\sigma}(x_i; \beta), \widetilde{\sigma}(x_i; \beta))$  will be referred to as  $\widetilde{F}_{\varepsilon(\beta)}$  within the binding function and its density will be denoted as  $\widetilde{f}_{\varepsilon(\beta)}$ . The corresponding quantities for the response variable are  $\widetilde{F}_{\widetilde{y}(\beta)}$  and  $\widetilde{f}_{\widetilde{y}(\beta)}$ .

To define a feasible indirect inference estimator, the simulated error distribution defined by  $\widetilde{\sigma}(x_i; \beta)$  has to be estimated by  $N(\mu_\tau \cdot \widehat{\sigma}_n(x_i; \beta^0), \widehat{\sigma}_n(x_i; \beta^0))$  using an estimate  $\widehat{\sigma}_n(x_i; \beta)$ . Denoting  $\widehat{\theta}_n$  the linear QR estimate for the original data and  $\widehat{F}_{y,n}(\cdot|x_i)$  an estimate of  $F_y(\cdot|x_i)$ , we define  $\widehat{\sigma}_n(x_i; \beta)$  as

$$\widehat{\sigma}_n(x_i; \beta) = \begin{cases} \frac{x_i^T \widehat{\theta}_n - x_i^T \beta}{\Phi_\tau^{-1}(\widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i))} & \text{if } x_i^T \beta \leq 0, \\ \frac{x_i^T \widehat{\theta}_n - x_i^T \beta}{\Phi_\tau^{-1}(\min\{\max\{\widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i) - \widehat{F}_{y,n}(x_i^T \beta | x_i) + \tau, 0\}, 1\})} & \text{if } x_i^T \beta > 0. \end{cases} \quad (17)$$

Since the denominators in (17) might take value 0, we again extend the definition (17) of  $\widehat{\sigma}_n(x_i; \beta)$  in such a way that the variance function  $\widehat{\sigma}_n(x_i; \beta)$  is continuous in  $x_i$ . For a given  $\beta$  and any sequence  $\{c_n\}_{n=1}^\infty$  such that  $c_n = O(n^{-k_0})$ ,  $k_0 > 0$ , suppose that  $|\widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i) - \widehat{F}_{y,n}(x_i^T \beta | x_i)| < c_n$  for  $x_i^T \beta > 0$  or  $|\widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i) - \tau| < c_n$  for  $x_i^T \beta \leq 0$ ; we refer to this event as the “zero-denominator”  $ZD_{i,n}(\widehat{\theta}_n, \beta)$ . If  $ZD_{i,n}(\widehat{\theta}_n, \beta)$  occurs, then we use instead of (17) the linearly interpolated values

$$\widehat{\sigma}_n(x_i; \beta) = \frac{x_i^T \widehat{\theta}_n - x_m^T \widehat{\theta}_n}{x_M^T \widehat{\theta}_n - x_m^T \widehat{\theta}_n} \widehat{\sigma}_n(x_M; \beta) + \frac{x_M^T \widehat{\theta}_n - x_i^T \widehat{\theta}_n}{x_M^T \widehat{\theta}_n - x_m^T \widehat{\theta}_n} \widehat{\sigma}_n(x_m; \beta), \quad (18)$$

where  $m = \arg \max_{j \leq n} \{x_j^T \widehat{\theta}_n : x_j^T \widehat{\theta}_n < x_i^T \widehat{\theta}_n \text{ and } I(ZD_{j,n}(\widehat{\theta}_n, \beta)) = 0\}$  and  $M = \arg \min_{j \leq n} \{x_j^T \widehat{\theta}_n : x_j^T \widehat{\theta}_n > x_i^T \widehat{\theta}_n \text{ and } I(ZD_{j,n}(\widehat{\theta}_n, \beta)) = 0\}$ ; if  $m = \emptyset$  or  $M = \emptyset$  (e.g., if  $\widehat{\theta}_n = \beta$ ),  $\widehat{\sigma}_n(x_m; \beta) = 1$  or  $\widehat{\sigma}_n(x_M; \beta) = 1$ , respectively. Alternatively, one can also use a straightforward analog of (15) and define

$$\widehat{\sigma}_n(x_i; \beta) = \frac{\phi_\tau \left\{ \Phi_\tau^{-1} \left[ \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i) \right] \right\}}{\widehat{f}_{y,n}(x_i^T \widehat{\theta}_n | x_i)}, \quad (19)$$

where  $\widehat{f}_{y,n}(\cdot|x_i)$  is an estimate of the conditional density function  $f_y(\cdot|x_i)$ . As this requires an additional nonparametric estimator, we rely on definition (18). The corresponding theoretical results in Sect. 4 are however valid also for (19) if the uniform convergence of  $\widehat{f}_{y,n}(\cdot|x_i)$  to  $f_y(\cdot|x_i)$  is imposed.

Having an estimate  $\widehat{\sigma}_n(x_i; \beta)$  defined by (17)–(18) (or (17) and (19)), the feasible indirect inference (FII) estimator  $\widehat{\beta}_n^{FII}$  can be defined as

$$\widehat{\beta}_n^{FII} = \arg \min_{\beta \in B} \left[ \widehat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \widehat{\theta}_n^s(\beta) \right]^T \Omega \left[ \widehat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \widehat{\theta}_n^s(\beta) \right], \quad (20)$$

where  $\widehat{\theta}_n^s(\beta) = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho_\tau(\widehat{y}_i^s(\beta) - x_i^T \theta)$  using  $\widehat{y}_i^s(\beta) = \max\{0, x_i^T \beta + \widehat{\varepsilon}_i^s\}$  and  $\widehat{\varepsilon}_i^s \sim N(\mu_\tau \cdot \widehat{\sigma}_n(x_i; \beta), \widehat{\sigma}_n(x_i; \beta))$ .

Finally, let us remark that the proposed estimator  $\widehat{\beta}_n^{FII}$  does not perform a selection procedure as it is done in [Khan and Powell \(2001\)](#) and [Chernozhukov and Hong \(2002\)](#), that is, the proposed estimation method is applied to all observations in the sample. Furthermore, our estimation procedure corrects the downward bias of linear QR caused by the censoring of the dependent variable and can be thus considered as a bias-correction method. The bias-correction procedure is based, similar to two-step estimators, on nonparametric estimates. Even though the bias-correction does not seem to be overly sensitive to the (lack of) precision of these nonparametric estimates, it could benefit from using some dimension reduction technique (e.g., [Xia et al. 2002](#)) to estimate  $\sigma(x_i, \beta)$  on a lower dimensional space in models with a large numbers of explanatory variables, especially discrete ones. Finally, the linear QR includes implicitly also quadratic and polynomial models since  $x_i$  can contain both the values of covariates as well as their powers. In the simulations (see Sect. 5), we used for simplicity only the auxiliary model linear in variables as the quadratic auxiliary model, which can better approximate the true quantile function, did not perform substantially better than the linear one.

## 4 Large sample properties

In this section, the asymptotic properties of the indirect-inference estimators for the censored regression model,  $\widehat{\beta}_n^{III}$  and  $\widehat{\beta}_n^{FII}$ , are derived. As our main result, we prove that  $\widehat{\beta}_n^{III}$  and  $\widehat{\beta}_n^{FII}$  are asymptotically equivalent and asymptotically normally distributed. Let us first introduce conditions required for establishing the consistency and asymptotic normality of the III estimator.

- A.1 The parameter spaces  $\Theta$  and  $B$  are compact subsets of  $R^k$  and the true parameter values are  $\theta^0 \in \Theta^\circ$  and  $\beta^0 \in B^\circ$ .
- A.2 The parameter vector  $\theta^0$  uniquely minimizes  $E[\rho_\tau(y_i - x_i^T \theta)]$ .
- A.3 The random vectors  $\{(x_i, y_i)\}_{i=1}^n$  are independent and identically distributed with finite second moments. The support of  $x_i \in X$  is assumed to be compact. Moreover, the index  $x_i^T \theta^0$  is continuously distributed, that is, there is at least one continuously distributed explanatory variable with  $\theta_j^0 \neq 0$ .
- A.4 The  $\tau$ th conditional quantile of  $\varepsilon_i$  is zero. The error term  $\varepsilon_i$  has the conditional distribution  $F_\varepsilon(t|x_i)$  with the conditional density function  $f_\varepsilon(t|x_i)$ , which is uniformly bounded both in  $t$  and  $x_i$ , positive on its support, and uniformly continuous with respect to  $t$  and  $x_i$ .

A.5 The following matrices are assumed to be finite and positive definite:

$$\begin{aligned} & - J_{crq} = E[I(x_i^T \beta^0 > 0) f_y(x_i^T \beta^0 | x_i) x_i x_i^T], \\ & - J = E[I(x_i^T \theta^0 > 0) f_{\tilde{y}}(x_i^T \theta^0 | x_i) x_i x_i^T], \\ & - \tilde{J} = E[I(x_i^T \theta^0 > 0) \tilde{f}_{\tilde{y}(\beta^0)}(x_i^T \theta^0 | x_i) x_i x_i^T], \\ & - \Sigma = E[\{\tau - I(y_i < x_i^T \theta^0)\}^2 x_i x_i^T], \\ & - \tilde{\Sigma} = E[\{\tau - I(\tilde{y}_i^s(\beta^0) < x_i^T \theta^0)\}^2 x_i x_i^T], \text{ and} \\ & - K = E[\{F_y(x_i^T \theta^0 | x_i) - \tau\}^2 x_i x_i^T]. \end{aligned}$$

A.6 Denoting  $\tilde{F}^0 = \tilde{F}_{\tilde{y}(\beta^0)}$ , the link function  $b(\tilde{F}^0, \beta)$  is a one-to-one mapping. Moreover,  $b(F, \beta)$  is assumed to be continuous in  $\beta$  and  $F$  (with respect to the supremum norm) at  $\beta^0$  and  $\tilde{F}^0$ . Finally,  $b(\tilde{F}_{\tilde{y}(\beta)}, \beta)$  is continuously differentiable in  $\beta \in U(\beta^0, \delta_b)$ ,  $\delta_b > 0$ , and  $D = \partial b(\tilde{F}_{\tilde{y}(\beta^0)}, \beta^0) / \partial \beta^T$  has a full column rank.

A.7  $P(\Delta(x_i, \theta^0), \beta^0) = 0$  and  $P(x_i^T \theta^0 = v) = 0$  for any  $v \in R$ .

Let us provide a few remarks regarding the necessity of these assumptions. Assumptions A.1, A.2, and A.3 are essential for establishing the consistency and asymptotic normality of the QR estimates  $\hat{\theta}_n$  as argued in [Angrist et al. \(2006\)](#) as well as the consistency and asymptotic normality of  $\hat{\beta}_n^{III}$ . As shown in [Angrist et al. \(2006\)](#), compactness of the support of  $X$ , which is typically achieved by trimming in semiparametric estimation and is also required by [Khan and Powell \(2001\)](#), can be relaxed to the existence of finite  $(2 + \delta)$ th moment of  $x_i$ . In our proofs of the asymptotic properties of  $\hat{\beta}_n^{III}$ , relaxing the compactness of  $X$  would additionally require  $\max_{i \leq n} \|x_i\| = o_p(n^\alpha)$  for some  $0 < \alpha < 1/2$  (this is an assumption closely related to the existence of finite  $(2 + \delta)$ th moments; see [Čížek 2006](#), Proposition 2.1). Moreover, we assume the existence of one continuous explanatory variable.

Next, Assumption A.4 is the standard assumption in quantile regression models (e.g., [Powell 1986a](#)), although the density function  $f_\varepsilon(t|x_i)$  is usually assumed to be positive only in a neighborhood of 0. Given the misspecification of the linear QR, it is convenient to assume non-zero density everywhere as  $f_\varepsilon(t|x_i)$  is evaluated for any  $t = x_i^T \theta^0$ . Concerning Assumption A.5, it contains usual full-rank conditions used in censored and quantile regression models and is necessary for the identification of parameter vectors, see for example [Khan and Powell \(2001\)](#). In the case of  $J$ ,  $\tilde{J}$ , and  $J_{crq}$ , it rules out collinearity among the explanatory variables restricted to regions with  $x_i^T \theta^0 > 0$  and  $x_i^T \beta^0 > 0$ , respectively (e.g., collinearity could arise if  $x_i$  with a bounded support contains a dummy variable  $d_i$  with so small coefficient that  $x_i^T \beta^0 < 0$  whenever  $d_i = 1$ ). Further, the first part of Assumption A.7 is imposed to simplify the proof of the consistency and asymptotic normality of the proposed estimator: it rules out the data without any censoring. The results remain valid even if there is no censoring, although some proofs would slightly differ. The second part of Assumption A.7 just formalizes the continuous-regressor Assumption A.3.

Finally, Assumption A.6 is the standard assumption necessary for defining the indirect inference estimator: the population QR estimates  $\tilde{\theta}(\beta)$  and  $\tilde{\theta}'(\beta')$  for data simulated from the censored regression model with parameters  $\beta$  and  $\beta'$  should differ if  $\beta \neq \beta'$ . Note though that we require the link function to be one-to-one only at the distribution  $\tilde{F}^0 = \tilde{F}_{\tilde{y}(\beta^0)}$  corresponding to the true parameter values  $\beta^0$ . This should however not be very restrictive in practice. On the one hand, the uniqueness of the

population QR estimates  $\theta^0$  and  $\tilde{\theta}(\beta)$  for some  $\beta$  requires, apart from the full-rank conditions in A.5, that the distribution of the responses and linear regression function is continuous with strictly positive conditional densities at  $x_i^T \theta^0$  for  $x_i^T \theta^0 > 0$  (cf. Assumption A.5). For the simulated data  $(x_i^T \theta^0, \tilde{y}_i^s(\beta))$ , the continuity follows from the presence of at least one continuously distributed explanatory variable (Assumption A.3 and A.7) and from the latent errors  $\varepsilon_i$  being Gaussian with uniformly bounded variances (see Appendix 2), which follows primarily from the compactness of the parameter and variable spaces (Assumptions A.1 and A.3). For the real data, which follow an unknown distribution, we however have to impose the uniqueness of  $\theta^0$  (Assumption A.2). On the other hand, the one-to-one link function also means that any change in  $\beta$  results in a change of  $\tilde{\theta}(\beta)$ . This holds trivially in the model without any censoring as the data generating process defined by  $\beta$  then leads to the population QR estimates  $\theta = b(\tilde{F}^0, \beta) = \beta$ . If the censoring is present,  $b(\tilde{F}^0, \beta) \neq \beta$ , but the link function is still usually one-to-one: any change in  $\beta$ , for example an increase of the intercept, is reflected by the corresponding change in the censored population data, for example by shifting non-censored population data up. This is a direct consequence of the error distribution defining  $\tilde{F}^0$  being independent of  $\beta$  and its conditional Gaussian density of  $\varepsilon_i$  given  $x_i$  being everywhere positive—the population data therefore always contain some non-censored responses at any  $x_i$ .

Although it is difficult to provide a more specific general condition which ensures that the link function is a bijection, one can detect the violation of this condition at least locally. Due to the local identification condition under model misspecification, which is represented here by the full rank of matrix  $D$  (cf. Theorem 3.1 of White 1982), singularity or near singularity of the estimated derivative of the link function will indicate that Assumption A.6 is violated. If one suspects that Assumption A.6 is violated, a possible approach is to enhance the auxiliary model in such a way that it better approximates the true model (the closer the auxiliary model is to the true model, the less likely Assumption A.6 is violated). In particular for the censored regression model, we can use, for example, quadratic functions of regressors and fit an auxiliary quadratic QR model rather than the linear one.

These assumptions are sufficient to derive the asymptotic distribution of the infeasible estimator. For the sake of simplicity of some proofs, we will additionally assume that the conditional error distribution  $F_\varepsilon(\cdot|x_i)$  has an infinite support (see Appendix 1 for details), but the stated results are valid in the general case as well.

**Theorem 2** *Let quantile  $\tau \in (0, 1)$ ,  $\Omega$  be a non-singular  $k \times k$  matrix, and  $S \in N$  be a fixed number of simulated samples. Under Assumptions A.1–A.7,  $\hat{\beta}_n^{III}$  is a consistent estimator of  $\beta^0$  and it is asymptotically normal:*

$$\sqrt{n}(\hat{\beta}_n^{III} - \beta^0) \rightarrow N\left(0, D^{-1}V(S)(D^T)^{-1}\right) \quad (21)$$

as  $n \rightarrow +\infty$ , where  $V(S) = J^{-1}\Sigma J^{-1} + \frac{1}{S}\tilde{J}^{-1}\tilde{\Sigma}\tilde{J}^{-1} + (1 - \frac{1}{S})\tilde{J}^{-1}K\tilde{J}^{-1} - 2J^{-1}K\tilde{J}^{-1}$ .

*Proof* See Appendix 3.  $\square$

The asymptotic variance matrix of  $\hat{\beta}_n^{III}$  derived in Theorem 2 consists of several parts. First, the matrices  $\Sigma$  and  $\tilde{\Sigma}$  are the variances of the QR first-order conditions in the real and simulated data, respectively. Next,  $J$  and  $\tilde{J}$  are the corresponding Jacobian matrices defined in Assumption A.5. Finally, matrix  $K$  characterizes the unconditional covariance between the real and simulated data.

The next theorem shows that the feasible estimator  $\hat{\beta}_n^{FII}$  is asymptotically equivalent to the infeasible one  $\hat{\beta}_n^{III}$  provided that one extra assumption holds: the conditional distribution function and its nonparametric estimates, which are used in (17) to define  $\hat{\sigma}_n(x_i; \beta)$ , have to be smooth functions of the data. Additionally, the nonparametric estimate  $\hat{F}_{y,n}(z_i|x_i)$  has to be consistent and to converge at a faster rate than the sequence  $c_n = O(n^{-k_0})$ ,  $k_0 > 0$ , used in the definition of  $\hat{\sigma}_n(x_i; \beta)$ . This is however not a constraint as  $k_0$  is arbitrary.

A.8 For any compact sets  $C_x \subset R^k$  and  $C_t \subset R^+$ ,  $\sup_{x \in C_x} \sup_{t \in C_t} |\hat{F}_{y,n}(t|x_i) - F_y(t|x_i)| = O_p(n^{-k_1})$  for some  $k_1 > k_0 > 0$  and  $\sup_{x \in C_x} \sup_{t \in C_t} |E\{\hat{F}_{y,n}(t|x_i) - F_y(t|x_i)\}| = o_p(n^{-1/2})$ . Moreover,  $\hat{F}_{y,n}(t|x_i)$  is monotonic in  $t$  and  $\hat{F}_{y,n}(t|x_i) = 0$  for  $t < 0$  and any  $x_i \in X$ .

A.9 The conditional distribution functions  $F_y(z|x_i = x)$  are piecewise Lipschitz functions in  $x$  for any  $z \in R$ .

Assumption A.8 is satisfied for many bias-corrected estimators of conditional distribution functions. Assumption A.9 on the conditional distribution function then states explicitly a minimum requirement that facilitates a consistent estimation and hence validity of Assumption A.8, although stronger assumptions on the smoothness of  $F_y(z|x_i)$  are usually used (cf. Li and Racine 2008).

**Theorem 3** *Let the assumptions of Theorem 2 be satisfied. If Assumptions A.8–A.9 also hold,  $\sqrt{n}(\hat{\beta}_n^{FII} - \hat{\beta}_n^{III}) \rightarrow 0$  in probability as  $n \rightarrow +\infty$ .*

*Proof* See Appendix 3.  $\square$

Theorem 3 shows that the feasible and infeasible II estimates,  $\hat{\beta}_n^{FII}$  and  $\hat{\beta}_n^{III}$ , are asymptotically equivalent, and consequently, the asymptotic variance-covariance matrix of  $\hat{\beta}_n^{FII}$  is given by (21). All elements of matrix  $V(S)$  can be readily estimated in practice (after replacing  $\theta^0$  by  $\hat{\theta}_n$ ) as we explain now. First, let us note that, as a by-product of constructing estimate  $\hat{\sigma}_n(x_i, \hat{\beta}_n^{FII})$  of  $\tilde{\sigma}(x_i, \beta^0)$  in the FII estimation procedure, we obtain along with the FII estimate  $\hat{\beta}_n^{FII}$  also estimates  $\hat{\theta}_n$ ,  $\hat{F}_{y,n}(x_i^T \hat{\theta}_n|x_i)$ , and  $\hat{f}_{y,n}(x_i^T \hat{\theta}_n|x_i)$  of  $\theta^0$ ,  $F_y(x_i^T \theta^0|x_i)$ , and  $f_y(x_i^T \theta^0|x_i)$ , respectively, for all  $i = 1, \dots, n$ ; see formulas (18) and (19). Moreover, all these estimates are consistent by Lemmas 5, 2, and 3. Another by-product of the estimation procedure are  $S$  simulated samples  $(x_i, \tilde{y}_i^s(\hat{\beta}_n^{FII}))_{i=1}^n$  used to compute  $S$  auxiliary estimates  $\hat{\theta}_n^s(\hat{\beta}_n^{FII})$  in (20). Since random vectors  $(x_i, y_i)$  are independent and identically distributed,  $i = 1, \dots, n$ , most elements of the asymptotic variance matrix  $D^{-1}[J^{-1}\Sigma J^{-1} + S^{-1}\tilde{J}^{-1}\tilde{\Sigma}\tilde{J}^{-1} + (1 - S^{-1})\tilde{J}^{-1}\tilde{K}\tilde{J}^{-1} - 2J^{-1}\tilde{K}\tilde{J}^{-1}](D^T)^{-1}$  can be consistently estimated due to the law of large numbers by (cf. Assumption A.5)

$$- \hat{J}_n = n^{-1} \sum_{i=1}^n I(x_i^T \hat{\theta}_n > 0) \hat{f}_{y,n}(x_i^T \hat{\theta}_n|x_i) x_i x_i^T;$$

- $\widehat{J}_n = n^{-1} \sum_{i=1}^n I(x_i^T \widehat{\theta}_n > 0) \widetilde{f}_{\widetilde{y}(\beta^0)}(x_i^T \widehat{\theta}_n | x_i) x_i x_i^T$ , where  $\widetilde{f}_{\widetilde{y}(\beta^0)}(\cdot | x_i)$  denotes the density of  $N(\mu_\tau, \widehat{\sigma}_n(x_i, \widehat{\beta}_n^{FII}))$ ;
- $\widehat{\Sigma}_n = n^{-1} \sum_{i=1}^n \{\tau - I(y_i < x_i^T \widehat{\theta}_n)\}^2 x_i x_i^T$ ;
- $\widehat{\widetilde{\Sigma}}_n = n^{-1} S^{-1} \sum_{i=1}^n \sum_{s=1}^S \{\tau - I(\widetilde{y}_i^s(\widehat{\beta}_n^{FII}) < x_i^T \widehat{\theta}_n)\}^2 x_i x_i^T$ , and
- $\widehat{K}_n = n^{-1} \sum_{i=1}^n \{\widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x_i) - \tau\}^2 x_i x_i^T$ .

The only exception is the matrix  $D = \partial b(\widetilde{F}_{\widetilde{y}(\beta^0)}, \beta^0) / \partial \beta^T$ , which represents the derivative of the link function. As this derivative is defined in terms of the simulated model with  $\widetilde{y}_i^s(\beta)$  being drawn from  $N(x_i^T \beta, \widehat{\sigma}_n(x_i, \beta))$  given  $x_i$  and censored at 0, we can draw an arbitrarily large sample representing the whole population and evaluate the link function and its derivative numerically as suggested by [Gouriéroux et al. \(1993\)](#). In particular, let  $1 \gg \delta > 0$ , an integer  $N \gg n$ , and  $\tilde{x}_k$  and  $\tilde{\varepsilon}_k$  be randomly drawn from the empirical distribution of  $\{x_i\}_{i=1}^n$  and  $N(0, 1)$  for  $k = 1, \dots, N$ . Denoting  $\widehat{\theta}_N(\beta)$  the auxiliary QR estimate obtained for data  $(\tilde{x}_k, \max\{0, \tilde{x}_k^T \beta + \tilde{\varepsilon}_k \widehat{\sigma}_n(\tilde{x}_k, \beta)\})_{k=1}^N$ , the derivative of the link function can be estimated by matrix  $\widehat{D}_n = (\widehat{d}_{ij,n})_{i,j=1}^{k,k}$  with the  $ij$ th element equal to

$$\widehat{d}_{ij,n} = \frac{\widehat{\theta}_{i,N}(\beta + \delta e_j) - \widehat{\theta}_{i,N}(\beta)}{\delta},$$

where  $\widehat{\theta}_{i,N}(\beta)$  denotes the  $i$ th element of  $\widehat{\theta}_N(\beta)$  and  $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$  is the vector with the  $j$ th element equal to 1 and all other elements equal to 0. Note though that, given the discussion in Sect. 5.2, it can be advisable to use a better approximation of the estimator's variance in small samples than the asymptotic distribution; for example, the bootstrap.

## 5 Monte Carlo simulations

Although we characterized the asymptotic properties of the proposed bias-corrected QR estimator, it is primarily aimed to improve the finite-sample performance of existing estimators. To analyze the benefits of the bias-correction performed by means of the indirect inference, this method is now compared with many existing estimators by means of Monte Carlo simulations. The simulation setting is described in Sect. 5.1 and the results are discussed in Sect. 5.2.

### 5.1 Simulation design

The data-generating process is similar to the one considered by [Khan and Powell \(2001\)](#):

$$y_i = \max\{\alpha + x_{1i} + \dots + x_{ki} + \varepsilon_i, 0\}$$

with the slope parameters equal to 1 and  $\alpha$  chosen in each sample so that the censoring level is always the same; unless stated otherwise, the censoring level equals 50%.



The  $k$  regressors  $x_{1i}, \dots, x_{ki}$  are uniformly distributed on  $\left(-\sqrt{3}, \sqrt{3}\right)$ , but the results are qualitatively rather similar across other data distributions (e.g.,  $x_i$  being normally distributed). Further, we focus on the median regression case  $\tau = 0.5$ . The error term  $\varepsilon_i$  can thus follow various error distributions with the median equal to zero, such as the normal  $N(0, \sigma_x)$ , Student  $t_d$ , and double exponential  $DExp(\lambda)$  ones.

For this data-generating process, we consider the following estimators: (i) the standard Tobit MLE constructed for normal homoscedastic errors; (ii) the CLAD estimator; (iii) the two-step LAD of [Khan and Powell \(2001\)](#) based on their three initial estimators—the maximum score estimator (2S-MS), the Nadaraya-Watson estimator of the propensity score (2S-NW), and the conditional quantile estimator (2S-LQR); (iv) the ‘infeasible’ LAD (IFLAD) defined as the QR estimator applied only to data points with  $\alpha + \beta x_i \geq 0$ ; (v) the three-step estimator of [Chernozhukov and Hong \(2002\)](#) based on the initial logit estimator (3S-LOG); (vi) the proposed QR estimator with bias corrected by indirect inference (FII); and (vii) the corresponding infeasible indirect inference (III) estimator, which does not estimate the conditional error distribution, but ‘knows’ the true one.

The QR estimates were in all cases computed by the Barrodale and Roberts (BR) algorithm as implemented in the R package “quantreg.” The same package was also used for computing CLAD by an adapted BR algorithm; given that it only finds local minima, we searched the global minimum by exhaustive search of all elemental subsets if  $k = 1$ , and due to infeasibility of this in higher dimensional models, started the adapted BR algorithm from the naive QR regression estimate if  $k \geq 1$ . For the indirect-inference based methods, we use the Nelder-Mead simplex method with multiple starting points as an optimization algorithm;<sup>2</sup> the number of simulated samples is  $S = 50$  by default. Further, many of the considered methods depend on some initial nonparametric estimators of the conditional mean, conditional quantile, and conditional distribution and density functions. The nonparametric estimators considered here are those by [Racine and Li \(2004\)](#) for the conditional mean and distribution function estimation, by [Li and Racine \(2008\)](#) for the conditional quantile estimation, and by [Hall et al. \(2004\)](#) for the conditional density estimation; we use their implementation in the R package “npreg,” which also includes the bandwidth choice by the least-squares cross-validation. The estimation and the bandwidth choice were based in all cases on the Gaussian kernel (the results are however insensitive to the kernel choice).

Finally, the bias-corrected QR estimator using the estimated values of the conditional distribution function can sometimes exhibit multiple minima in very small samples (in such cases, there are usually two minima found irrespective of the number of starting points). This concerns only the feasible estimator, but not the infeasible one: given the consistency of the nonparametric estimators, of the proposed II method, and the identification result in [Lemma 1](#), multiple minima thus disappear as the sample size increases (as we practically observed in the simulation study). In the presence of multiple local minima, we simply use the average of the found local minima as the

<sup>2</sup> Note that multiple starting points are relevant only for the feasible II estimator in small samples as discussed later. The III method does not exhibit multiple minima at any sample size.



**Table 1** The bias and root mean squared error of the indirect inference estimators based on  $S$  simulated samples in a model with the standard normal errors

Estimator	III		FII	
	Bias	RMSE	Bias	RMSE
$S$				
5	0.001	0.189	−0.086	0.252
10	0.005	0.187	−0.084	0.255
25	0.009	0.191	−0.071	0.267
50	0.008	0.185	−0.065	0.278
100	0.010	0.188	−0.057	0.286
250	0.008	0.186	−0.051	0.297

estimate. Unreported simulations indicate that multiple minima in small samples do not occur at all if the quadratic auxiliary model is used.

The results for all methods are obtained using 1000 simulations for  $k \in \{1, 3, 5\}$  variables and sample sizes  $n = 50, 100$ , and  $200$  and are summarized using the bias and root mean squared error (RMSE) of the slope estimates (the absolute bias is reported if  $k > 1$ ).

## 5.2 Simulation results

Let us first discuss the number  $S$  of simulated samples and its influence on the estimates for standard normal errors,  $k = 1$ , and  $n = 100$ . As Table 1 documents, the II estimators are consistent for any fixed and finite  $S$ : the III estimator performs equally well for any  $S = 5, \dots, 250$  and one could expect similar performance of FII in large samples once the bias of nonparametric estimation becomes small. At  $n = 100$  however, FII based on an increasing number  $S$  of simulated samples shows a diminishing negative finite-sample bias and an increasing RMSE (RMSE does not increase further for  $S > 250$ ). While the trade-off is not too large and we choose  $S = 50$  here,  $S$  should be set larger if one is concerned about the bias and can be set smaller if RMSE is important. A similar trade-off—a larger negative bias connected to a lower variance of estimates—will be also observed later in the case of some existing estimators.

The first comparison of estimators is obtained for  $\varepsilon_i \sim N(0, 1)$ ,  $k = 1$ , and  $n = 50, 100$ , and  $200$ ; see Table 2. The MLE estimator serves as a parametric benchmark, and given normality, performs best in all cases. First, CLAD exhibit large biases and RMSEs in small samples, especially with  $n = 50$  observations; this is due to the heavy right-tail of the CLAD distribution. Next, the existing two-step estimators exhibit relatively large RMSEs, which are however smaller than those of CLAD (except for 2S-MS), and negative biases, which vary with the choice of the initial estimator. The fact that 2S-LQR performs better than the infeasible IFLAD in terms of RMSE is related to the difference in definitions: IFLAD uses data points with  $\alpha + \beta x_i \geq 0$ , whereas the two-step estimators rely on data points with  $\alpha + \beta x_i \geq c$  (e.g.,  $c$  is set to 0.05 here as done in Khan and Powell 2001). Increasing  $c$  reduces the bias, but increases the variance of estimates (especially for 2S-LQR) and vice versa. In comparison, the three-step estimator 3S-LOG has usually larger RMSEs than 2S-NW

**Table 2** The bias and root mean squared error of all estimators in a model with the standard normal errors at different sample sizes  $n = 100, 200$ , and  $400$ 

Estimator	$n = 50$		$n = 100$		$n = 200$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	-0.010	0.199	-0.004	0.138	-0.001	0.097
CLAD	0.402	1.228	0.140	0.481	0.078	0.292
<i>IFLAD</i>	<i>-0.109</i>	<i>0.438</i>	<i>-0.090</i>	<i>0.298</i>	<i>-0.061</i>	<i>0.204</i>
2S-MSc	-0.182	0.641	-0.196	0.506	-0.124	0.388
2S-NW	-0.114	0.538	-0.103	0.396	-0.060	0.269
2S-LQR	-0.146	0.412	-0.121	0.294	-0.084	0.220
3S-LOG	-0.050	0.506	-0.049	0.373	-0.028	0.267
<i>III</i>	<i>0.037</i>	<i>0.276</i>	<i>0.008</i>	<i>0.191</i>	<i>0.005</i>	<i>0.132</i>
		(0.303)		(0.214)		(0.151)
FII	-0.040	0.398	-0.066	0.276	-0.051	0.218

The values in brackets represent the standard deviations based on the asymptotic distribution of the II estimators

The results for the infeasible estimators are in italics

or 2S-LQR, but rather small finite-sample bias compared to all other semiparametric methods.

Looking at the two infeasible estimators, III exhibits always smaller RMSE than IFLAD, although the difference decreases with an increasing sample size. It is also interesting to note that IFLAD exhibits systematically a larger negative bias, whereas III leads to a smaller, but positive bias (or almost zero bias for  $n = 200$ ). This is reflected by the performance of the proposed FII estimator, which exhibits smaller biases and RMSEs than any of the existing semiparametric methods. One can also notice that, even though FII exhibits generally a smaller bias than the methods of [Khan and Powell \(2001\)](#), the bias of FII is negative in contrast to the bias of III. Finally, let us mention the variance of the II estimates implied by the asymptotic distribution in Theorem 2. While the simulated and asymptotic RMSE are approaching each other in the case of III, the asymptotic standard deviations differ substantially in the case of the feasible estimator FII. Given that the small-sample variance of FII is additionally related to the number  $S$  of simulated samples in the opposite way than in Theorem 2, see Table 1, it is not advisable to use the asymptotic distribution in small samples.

The next set of results is obtained for five different error distributions, see Table 3 for  $k = 1$  and  $n = 100$ . The first three distributions are homoscedastic— $N(0, 1)$ ,  $t_5$ , and  $Exp(1)$ , whereas the remaining two distribution are heteroscedastic— $N(0, ce^{0.75x})$  in the case of “positive” heteroscedasticity and  $N(0, ce^{-0.75x})$  in the case of “negative” heteroscedasticity ( $c$  is always chosen so that the unconditional variance equals 1). The “Gaussian” MLE estimator serves again as a parametric benchmark, and in the case of homoscedasticity, performs very well across the tested (symmetric and unimodal) distributions. The performance of the semiparametric estimators under homoscedasticity does not substantially differ from the case with normal errors. Besides 2S-MSc, the semiparametric estimators possess smaller RMSEs than CLAD. The two-step esti-

**Table 3** The bias and root mean squared error of all estimators in a model with the standard normal, student, and double exponential errors;  $n = 100$ 

Estimator	$\varepsilon \sim N(0, 1)$		$\varepsilon \sim t_5$		$\varepsilon \sim DExp(1)$		$\varepsilon \sim N(0, e^{0.75x})$		$\varepsilon \sim N(0, e^{-0.75x})$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	-0.004	0.138	0.050	0.179	0.079	0.200	0.606	0.655	-0.445	0.458
CLAD	0.140	0.481	0.183	0.524	0.116	0.448	0.292	0.951	0.036	0.186
<i>IFLAD</i>	<i>-0.090</i>	<i>0.298</i>	<i>-0.081</i>	<i>0.296</i>	<i>-0.089</i>	<i>0.281</i>	<i>-0.048</i>	<i>0.334</i>	<i>-0.032</i>	<i>0.151</i>
2S-MS-C	-0.196	0.506	-0.158	0.517	-0.174	0.495	-0.110	0.604	-0.076	0.326
2S-NW	-0.103	0.396	-0.092	0.417	-0.096	0.360	-0.023	0.416	-0.062	0.202
2S-LQR	-0.121	0.294	-0.118	0.309	-0.133	0.304	-0.050	0.310	-0.120	0.241
3S-LOG	-0.049	0.373	-0.012	0.396	-0.045	0.351	0.037	0.451	-0.034	0.210
<i>III</i>	<i>0.008</i>	<i>0.191</i>	<i>0.031</i>	<i>0.224</i>	<i>0.034</i>	<i>0.239</i>	<i>0.013</i>	<i>0.265</i>	<i>0.023</i>	<i>0.185</i>
FII	-0.066	0.276	-0.034	0.304	-0.002	0.313	0.076	0.348	0.032	0.198

The results for the infeasible estimators are in italics

**Table 4** The absolute bias and root mean squared error of all estimators in a model with the standard normal errors and  $k = 1, 3$ , and 5 explanatory variables;  $n = 100$ 

Estimator	$k = 1$		$k = 3$		$k = 5$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	0.004	0.138	0.005	0.238	0.007	0.315
CLAD	0.140	0.481	0.070	0.434	0.034	0.502
<i>IFLAD</i>	<i>0.090</i>	<i>0.298</i>	<i>0.067</i>	<i>0.346</i>	<i>0.081</i>	<i>0.444</i>
2S-MS-C	0.196	0.506	0.156	0.486	0.160	0.534
2S-NW	0.103	0.396	0.336	0.513	0.488	0.643
2S-LQR	0.121	0.294	0.414	0.515	0.523	0.662
3S-LOG	0.049	0.373	0.091	0.403	0.131	0.486
<i>III</i>	<i>0.008</i>	<i>0.191</i>	<i>0.067</i>	<i>0.353</i>	<i>0.073</i>	<i>0.482</i>
FII	0.066	0.276	0.065	0.385	0.067	0.487

The results for the infeasible estimators are in italics

matrices however exhibit relatively large negative biases compared to the three step estimator and the FII estimator. The FII estimator is thus preferable in all cases.

The comparison substantially changes once the heteroscedastic errors are considered. The MLE estimates are now severely biased. In the “positive” heteroscedasticity scenario, CLAD exhibits large bias and RMSE. The existing two- and three-step estimators thus provide better estimates than CLAD in terms of RMSEs, and moreover, their biases increase (becoming less negative or even positive). The best performance can be attributed to 2S-LQR followed by FII. In the case of the “negative” heteroscedasticity, the differences among semiparametric estimators are much smaller and CLAD becomes the best performing estimator due to the fact that the observations with con-

ditional median above zero have now very low variance (cf. Khan and Powell 2001). Note that the proposed FII is again the second best, now after CLAD.

Furthermore, let us consider a model with standard normal errors and multiple regressors,  $k = 1, 3, 5$ ; in Table 4, the sample size  $n = 100$  is used irrespective of the dimension  $k$ . Let us first state that the results obtained for  $k = 1$  in previous experiments for CLAD and 2S-MSR are not fully comparable to the results for  $k = 3$  and  $k = 5$  as the earlier ones ( $k = 1$ ) search for global minima while the latter ones ( $k > 1$ ) correspond to the local minima obtained by starting from the linear QR estimates. This of course reduces variability of the corresponding estimates, and in this experimental design, makes CLAD and 2S-MSR competitive. Nevertheless, one can observe that the semiparametric methods based on some initial  $k$ -dimensional nonparametric smoothing tend to increase their RMSE with  $k$  faster than methods based purely on one-dimensional (non)parametric estimation. The large part of this effect is related to the increases in the total absolute bias—see the biases of 2S-NW and 2S-LQR, which are similar for each individual parameter irrespective of  $k$ , but the total bias of  $k$  estimated parameter thus grows with  $k$ . The II bias-correction procedure is not affected by this and FII thus performs substantially better than 2S-NW and 2S-LQR for larger  $k$ . On the other hand, the performance of FII is matched by the three-step method of Chernozhukov and Hong (2002) at  $k = 5$  and also by the “local-minimum” CLAD.

Finally, let us shortly mention how the estimators are affected by the amount of censored observations in the data. Using standard normal errors,  $k = 1$ , and  $n = 100$ , we studied performance under the assumption that 50 % observations are censored. If the fraction of censored observations decreases towards zero, the asymptotic theory predicts that all considered semiparametric estimates will converge to the simple linear QR estimates. This is confirmed by the simulation results in Table 5: we see that, at lower levels of censoring and in particular for 25 % censored observations, the RMSEs of all semiparametric estimators are very close to each other (with the exception of 2S-MSR).

**Table 5** The bias and root mean squared error of all estimators in a model with the standard normal errors and 25 %, 37.5, and 50 % censored observations;  $n = 100$

Estimator	25 %		37.5 %		50 %	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	−0.009	0.113	−0.005	0.125	−0.004	0.138
CLAD	0.037	0.192	0.050	0.262	0.140	0.481
<i>IFLAD</i>	<i>−0.041</i>	<i>0.153</i>	<i>−0.061</i>	<i>0.210</i>	<i>−0.090</i>	<i>0.298</i>
2S-MSR	−0.108	0.245	−0.152	0.340	−0.196	0.506
2S-NW	−0.093	0.185	−0.088	0.244	−0.103	0.396
2S-LQR	−0.114	0.183	−0.115	0.225	−0.121	0.294
3S-LOG	−0.025	0.189	−0.030	0.262	−0.049	0.373
<i>III</i>	<i>0.010</i>	<i>0.144</i>	<i>0.004</i>	<i>0.166</i>	<i>0.008</i>	<i>0.191</i>
FII	−0.011	0.186	−0.027	0.243	−0.066	0.276

The results for the infeasible estimators are in italics

Altogether, all semiparametric alternatives to CLAD perform better than CLAD, although the differences are likely to be small for very large data sets. The proposed FII estimator performs equally well in large samples and is in many cases preferable to existing semiparametric methods in small and moderate samples.

## 6 Conclusion

We proposed a new estimation method for the censored regression models that—contrary to existing methods—relies on the linear QR estimates for the whole sample and that applies a bias-correction technique to obtain consistent estimates. For the bias correction, the indirect inference technique is applied and extended so that it allows sampling from a nonparametrically estimated distribution function. The consistency and asymptotic distribution of the proposed estimator were found and shown to be first-order independent of the initial nonparametric estimates of the auxiliary error distribution. Finally, one of the important benefits of this estimation approach is its small-sample performance as was demonstrated by means of Monte Carlo simulations.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix 1: Proof of Theorem 1

*Proof of Theorem 1* The proof is almost identical to the proof of Theorem 2 in Angrist et al. (2006). We need to prove that the solution of

$$\theta^0 = \arg \min_{\theta \in \Theta} E \left[ \rho_{\tau}(y_i - x_i^T \theta) \right] \quad (22)$$

is equal to the solution of

$$\bar{\theta} = \arg \min_{\theta \in \Theta} E \left[ w(x_i, \beta^0, \bar{\theta}) \cdot \Delta^2(x_i, \beta^0, \theta) \right]. \quad (23)$$

Since the objective function in (23) is convex, any fixed point  $\theta = \bar{\theta}$  is a solution of the corresponding first-order condition:

$$\mathcal{F}(\theta) = 2 \cdot E \left[ w(x_i, \beta^0, \theta) \cdot \Delta(x_i, \beta^0, \theta) \cdot x_i \right] = 0. \quad (24)$$

On the other hand, the first order condition for (22) is given by (cf. the proof of Theorem 2 in Angrist et al. 2006)

$$\mathcal{D}(\theta) = E \left[ \left\{ I \left( u_i \leq \Delta(x_i, \beta^0, \theta) \right) - \tau \right\} \cdot x_i \right] = 0. \quad (25)$$

By the law of iterated expectations,  $D(\theta)$  can be written as

$$\mathcal{D}(\theta) = E \left[ \left\{ F_{u_i} \left( \Delta(x_i, \beta^0, \theta) | x_i \right) - \tau \right\} \cdot x_i \right] = 0. \quad (26)$$

Further, it follows from the definition of  $w(x_i, \beta^0, \theta)$  that  $F_{u_i}(\Delta(x_i, \beta^0, \theta) | x_i) - \tau = 2 \cdot w(x_i, \beta^0, \theta) \cdot \Delta(x_i, \beta^0, \theta)$  for any value of  $\Delta(x_i, \beta^0, \theta) \neq 0$ . As  $P(\Delta(x_i, \beta^0, \theta) = 0 | x_i) = 0$  and both  $F_{u_i}(\Delta(x_i, \beta^0, \theta) | x_i)$  and  $x_i$  are uniformly bounded by Assumption A.3, it holds that  $E[\{F_{u_i}(\Delta(x_i, \beta^0, \theta) | x_i) - \tau\} \cdot x_i \cdot I\{\Delta(x_i, \beta^0, \theta) = 0\}] = 0$ , and consequently,

$$\begin{aligned} \mathcal{D}(\theta) &= E \left[ \left\{ F_{u_i} \left( \Delta(x_i, \beta^0, \theta) | x_i \right) - \tau \right\} \cdot x_i \right] \\ &= E \left[ \left\{ F_{u_i} \left( \Delta(x_i, \beta^0, \theta) | x_i \right) - \tau \right\} \cdot x_i \cdot I \left\{ \Delta(x_i, \beta^0, \theta) \neq 0 \right\} \right] \\ &= 2 \cdot E \left[ w(x_i, \beta^0, \theta) \cdot \Delta(x_i, \beta^0, \theta) \cdot x_i \right] = \mathcal{F}(\theta). \end{aligned}$$

Because  $\theta^0$  is the unique solution of (22), it also uniquely solves (23) since the objective function in (23) is convex in  $\theta$ . Therefore,  $\theta = \theta^0 = \bar{\theta}$  solves both (22) and (23).  $\square$

## Appendix 2: Auxiliary lemmas

For the rest of the proofs, we introduce necessary notation. The norms  $\|\cdot\|$  and  $\|\cdot\|_\infty$  will refer to the Euclidean norm on  $R^d$  and to the supremum norm in functional spaces, respectively. The  $\delta$ -neighborhood of a vector  $t \in R^d$  is denoted  $U(t, \delta) = \{t' \in R^d : \|t' - t\| < \delta\}$ . The probability distribution and density functions of  $N(-\mu_\tau, 1)$  are denoted  $\Phi_\tau$  and  $\phi_\tau$ , respectively. Additionally, recall that  $\rho_\tau(z) = \{\tau - I(z \leq 0)\}z$  and its derivative is denoted  $\varphi_\tau(z) = \tau - I(z \leq 0)$ .

Next, for  $w_i = x_i$  or  $w_i = (y_i, x_i)$ , let  $E_n[f(w_i)]$  denote  $n^{-1} \sum_{i=1}^n f(w_i)$  and let  $G_n[f(w_i)]$  denote  $n^{-1/2} \sum_{i=1}^n \{f(w_i) - E[f(w_i)]\}$ . If we need to indicate a particular data distribution  $P$  of  $w_i$ ,  $G_{n,P}[f(w_i)] = n^{-1/2} \sum_{i=1}^n \{f(w_i) - E_P[f(w_i)]\}$  is used, assuming that  $w_i \sim P$ . For easier reading, we also use a simplified notation for the simulated distributions  $\tilde{F}(\beta) = \tilde{F}_{\tilde{y}(\beta)}$  and  $\hat{F}(\beta) = \hat{F}_{\hat{y}(\beta)}$ .

For the sake of simplicity of some proofs, we will additionally assume that the conditional error distribution  $F_\varepsilon(\cdot | x_i)$  has (uniformly) an infinite support in order to guarantee that  $\sup_{x \in X} F_\varepsilon(K | x) < 1$  for any  $K < \infty$ , and by Assumption A.3, that  $\sup_{x \in X} \sup_{\theta \in \Theta} F_y(x^T \theta | x) < K_F < 1$ . Consequently, the conditional variance  $\tilde{\sigma}(x_i; \beta^0)$  defined in (14)–(15) is everywhere positive at the true  $\beta^0$ , and given the compactness of  $B$ ,  $\Theta$ ,  $X$ , and A.7,  $\tilde{\sigma}(x_i; \beta^0) > C_\sigma > 0$  for all  $x_i \in X$ . If the limit expression (15) and Assumption A.4 are taken into account, one can observe that the variance function is also bounded from above:  $\tilde{\sigma}(x_i; \beta) < K_\sigma$  for any  $\beta \in B$  and all  $x_i \in X$ .

First, the following identification result is derived.

**Lemma 1** Under Assumptions A.1–A.5 and A.7,  $b(\tilde{F}(\beta^0), \beta^0) \neq b(\tilde{F}(\beta^1), \beta^1)$  for any  $\beta^1 \in B$  such that  $\beta^1 \neq \beta^0$ .

*Proof* First, note that under the listed assumptions,  $b(\tilde{F}^0, \beta)$  is one-to-one, where  $\tilde{F}^0 = \tilde{F}(\beta^0)$ . Suppose that  $b(\tilde{F}^0, \beta^0) = b(\tilde{F}(\beta^0), \beta^0) = b(\tilde{F}(\beta^1), \beta^1)$ . Since  $b(\tilde{F}^0, \beta)$  is one-to-one, the vector  $\beta^0$ , which satisfies  $b(\tilde{F}^0, \beta^0) = b(F_y, \beta^0) = \theta^0$  by Theorem 1, uniquely solves the QR moment condition

$$E[F_y(b(\tilde{F}^0, \beta^0)|x_i) - \tau] \cdot x_i = 0 \quad (27)$$

[see Eq. (26)]. As  $F_y(t|x_i) = 0$  for any  $t < 0$  and  $P(x_i^T \theta^0 = 0) = 0$  by Assumption A.7, (27) can be written as

$$\begin{aligned} & E[I(x_i^T b(F_y, \beta^0) > 0) \cdot \{F_y(x_i^T b(F_y, \beta^0)|x_i) - \tau\} \cdot x_i] \\ & = E[I(x_i^T b(F_y, \beta^0) < 0) \cdot \tau \cdot x_i] = k_0. \end{aligned} \quad (28)$$

Next, the QR moment condition for the simulated data  $\{\tilde{y}_i^s(\beta^1), x_i\}_{i=1}^n$  is [see Eqs. (25)–(26)]

$$E[(\tilde{F}_{\tilde{y}(\beta^1)}(x_i^T b\{\tilde{F}(\beta^1), \beta^1\}|x_i) - \tau) \cdot x_i] = 0. \quad (29)$$

Because the censored distribution  $\tilde{F}_{\tilde{y}(\beta^1)}(t|x_i) = 0$  for all  $t < 0$ , we can again rewrite it as

$$\begin{aligned} & E[I(x_i^T b\{\tilde{F}(\beta^1), \beta^1\} > 0) \cdot (\tilde{F}_{\tilde{y}(\beta^1)}(x_i^T \theta^0|x_i) - \tau) \cdot x_i] \\ & = E[I(x_i^T b\{\tilde{F}(\beta^1), \beta^1\} < 0) \tau \cdot x_i]. \end{aligned}$$

Recalling that  $\theta^0 = b(F_y, \beta^0) = b(\tilde{F}^0, \beta^0) = b(\tilde{F}(\beta^0), \beta^0) = b(\tilde{F}(\beta^1), \beta^1)$  and that  $\tilde{F}_{\tilde{y}(\beta)}(t|x_i) = \Phi_\tau\{(t - x_i^T \beta)\tilde{\sigma}^{-1}(x_i, \beta)|x_i\}$  for  $t > 0$ , (29) becomes

$$\begin{aligned} & E[I(x_i^T b(\tilde{F}^0, \beta^0) > 0) \cdot (\Phi_\tau\{[x_i^T b(\tilde{F}(\beta^1), \\ & \beta^1) - x_i^T \beta^1] \cdot \tilde{\sigma}^{-1}(x_i, \beta^1)|x_i\} - \tau) \cdot x_i] = k_0. \end{aligned}$$

By substituting (14),<sup>3</sup> where  $\theta^0$  is replaced by  $b(\tilde{F}^0, \beta^0)$ , we get

$$\begin{aligned} & E\left[I\left(x_i^T b\left(\tilde{F}^0, \beta^0\right) > 0\right) I\left(x_i^T \beta^1 \leq 0\right)\right. \\ & \quad \times \left\{\Phi_\tau\left(\frac{x_i^T b\left\{\tilde{F}\left(\beta^1\right), \beta^1\right\} - x_i^T \beta^1}{x_i^T b\left(\tilde{F}^0, \beta^0\right) - x_i^T \beta^1} \cdot \Phi_\tau^{-1}\left\{F_y\left(x_i^T b\left(\tilde{F}^0, \beta^0\right)|x_i\right)\right\}\right) - \tau\right\} x_i \\ & \quad \left.+ I\left(x_i^T b\left(\tilde{F}^0, \beta^0\right) > 0\right) I\left(x_i^T \beta^1 > 0\right)\right] \end{aligned}$$

<sup>3</sup> We drop min and max conditions in (14) since  $\sigma(x_i; \beta^0) = \tilde{\sigma}(x_i; \beta^0) > C_\sigma > 0$  by definition.

$$\times \left\{ \Phi_{\tau} \left( \frac{x_i^T b \left\{ \tilde{F}(\beta^1), \beta^1 \right\} - x_i^T \beta^1}{x_i^T b(\tilde{F}^0, \beta^0) - x_i^T \beta^1} \cdot \Phi_{\tau}^{-1} \left\{ F_y(x_i^T b(\tilde{F}^0, \beta^0)|x_i) - F_y(x_i^T \beta^1|x_i) + \tau \right\} \right) - \tau \right\} x_i \Bigg] = k_0.$$

Recalling again  $b(\tilde{F}^0, \beta^0) = b(\tilde{F}(\beta^0), \beta^0) = b(\tilde{F}(\beta^1), \beta^1)$ , we obtain

$$\begin{aligned} & E[I(x_i^T b(\tilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 \leq 0) \cdot \{F_y(x_i^T b(\tilde{F}^0, \beta^0)|x_i) - \tau\} \cdot x_i \\ & + I(x_i^T b(\tilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 > 0) \cdot \{F_y(x_i^T b(\tilde{F}^0, \beta^0)|x_i) \\ & - F_y(x_i^T \beta^1|x_i)\} \cdot x_i] = k_0. \end{aligned} \quad (30)$$

Using identity (28), (30) can be simplified to

$$E[I(x_i^T b(\tilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^1 > 0) \cdot (F_y(x_i^T \beta^1|x_i) - \tau) \cdot x_i] = 0. \quad (31)$$

By Powell (1986a), under Assumptions A.1–A.5 we know that  $E[I(x_i^T \beta^0 > 0) \cdot (F_y(x_i^T \beta^0|x_i) - \tau) \cdot x_i] = 0$  which implies

$$E[I(x_i^T b(\tilde{F}^0, \beta^0) > 0) \cdot I(x_i^T \beta^0 > 0) \cdot (F_y(x_i^T \beta^0|x_i) - \tau) \cdot x_i] = 0$$

since  $\{x_i \in R^k | I(x_i^T \beta^0 > 0) I(x_i^T b(\tilde{F}^0, \beta^0) > 0)\} \subset \{x_i \in R^k | I(x_i^T \beta^0 > 0)\}$ . Thus, we should have  $\beta^0 = \beta^1$  because  $J_{crq}$  is positive definite by Assumption A.5.  $\square$

The further auxiliary results presented in this section are proved in the discussion paper of Čížek and Sadikoglu (2014).

**Theorem 4** Let  $J = E[I(x_i^T \theta > 0) f_y(x_i^T \theta|x_i) x_i x_i^T]$ . Under Assumptions A.1–A.5 and A.7, it holds that

1.  $Q_n(\theta) = E_n[\rho_{\tau}(y_i - x_i^T \theta) - \rho_{\tau}(y_i - x_i^T \theta^0)] \rightarrow Q_{\infty}(\theta) = E[\rho_{\tau}(y_i - x_i^T \theta) - \rho_{\tau}(y_i - x_i^T \theta^0)]$  as  $n \rightarrow \infty$  for any  $\theta \in \Theta$ ;
2.  $\hat{\theta}_n$  is a consistent estimator of  $\theta^0$ ,  $\hat{\theta}_n \xrightarrow{P} \theta^0$  as  $n \rightarrow \infty$ ;
3. for any sequence  $\theta_n \xrightarrow{P} \theta^0$ ,  $n^{1/2}(\theta_n - \theta^0) = -J^{-1} G_n\{\varphi_{\tau}(y_i - x_i^T \theta^0) x_i\} + o_p(1)$  converges to a Gaussian process with covariance function  $\Sigma = E[(\tau - I(y_i < x_i^T \theta^0))(\tau - I(y_i < x_i^T \theta^0)) x_i x_i^T]$ .

Additionally, suppose that, for sample size  $n \in N$ , data are independently and identically distributed according to probability distributions  $P_n$ , which satisfy Assumptions A.1–A.5 and A.7 uniformly in  $n$ . Denoting

$$r(y, x, \theta) = \begin{cases} [\rho_{\tau}(y - x^T \theta) - \rho_{\tau}(y - x^T \theta^0) \\ - (\theta - \theta^0)^T (\tau - I\{y \leq x^T \theta^0\}) \cdot x] / \|\theta - \theta^0\| & \text{if } \theta \neq \theta^0, \\ 0 & \text{if } \theta = \theta^0, \end{cases}$$



let us assume that

$$\sup_{\theta', \theta'' \in U(\theta^0, \delta)} |E_{P_n}\{r(y_i, x_i, \theta') - r(y_i, x_i, \theta'')\}^2 - E_{P_0}\{r(y_i, x_i, \theta') - r(y_i, x_i, \theta'')\}^2| \rightarrow 0 \quad (32)$$

as  $n \rightarrow \infty$  for some  $\delta > 0$  and some distribution  $P_0$ , which satisfies assumptions A.1–A.5 and A.7. Then for any sequence  $\theta_n \xrightarrow{P} \theta^0$ ,  $n^{1/2}(\theta_n - \theta^0) = -J^{-1}G_{n, P_n}\{\varphi_\tau(y_i - x_i^T \theta^0)x_i\} + o_p(1)$  as  $n \rightarrow \infty$ .

*Proof* See Čížek and Sadikoglu (2014, Theorem 5).  $\square$

**Lemma 2** Under Assumptions A.1–A.5, A.7, and A.8, it holds for  $n \rightarrow \infty$  that

$$\sup_{x \in X} \left| \widehat{F}_{y,n}(x_i^T \widehat{\theta}_n | x) - F_y(x_i^T \theta^0 | x) \right| = O_p(n^{-\min\{k_1, 1/2\}}).$$

*Proof* See Čížek and Sadikoglu (2014, Lemma 7).  $\square$

**Lemma 3** Under Assumptions A.1–A.5, A.7, and A.8, it holds for any  $\beta \in B$ , sufficiently small  $\delta > 0$ ,  $l \in \{1, 2\}$ , and  $n \rightarrow \infty$  that

$$|\widehat{\sigma}_n(x_i; \beta) - \widetilde{\sigma}(x_i; \beta)| = o_p(1), \quad (33)$$

$$\sup_{\beta \in U(\beta^0, \delta)} |\widehat{\sigma}_n(x_i; \beta) - \widetilde{\sigma}(x_i; \beta)| = o_p(1), \quad (34)$$

$$E|\widehat{\sigma}_n(x_i; \beta) - \widetilde{\sigma}(x_i; \beta)|^l = o(1), \quad (35)$$

$$E\left(\sup_{\beta \in U(\beta^0, \delta)} |\widehat{\sigma}_n(x_i; \beta) - \widetilde{\sigma}(x_i; \beta)|\right)^l = o_p(1). \quad (36)$$

*Proof* See Čížek and Sadikoglu (2014, Lemma 8).  $\square$

**Corollary 1** Under Assumptions A.1–A.5, A.7, and A.8, it holds for some  $\delta > 0$  and  $n \rightarrow \infty$ :

$$\sup_{\beta \in U(\beta^0, \delta)} \|\widehat{F}(\beta) - \widetilde{F}(\beta)\|_\infty = \sup_{\beta \in U(\beta^0, \delta)} \sup_{t \in R} \left| \Phi_\tau \left[ \frac{t - x_i^T \beta}{\widehat{\sigma}_n(x_i; \beta)} \right] - \Phi_\tau \left[ \frac{t - x_i^T \beta}{\widetilde{\sigma}(x_i; \beta)} \right] \right| = o_p(1),$$

$$\sup_{\beta \in U(\beta^0, \delta)} \sup_{t \in R} \left| \frac{1}{\widehat{\sigma}_n(x_i; \beta)} \phi_\tau \left[ \frac{t - x_i^T \beta}{\widehat{\sigma}_n(x_i; \beta)} \right] - \frac{1}{\widetilde{\sigma}(x_i; \beta)} \phi_\tau \left[ \frac{t - x_i^T \beta}{\widetilde{\sigma}(x_i; \beta)} \right] \right| = o_p(1).$$

Further, function  $\widetilde{F}(\beta)$  is continuous in  $\beta$  on  $U(\beta^0, \delta)$ :

$$\lim_{\beta' \rightarrow \beta} \|\widetilde{F}(\beta') - \widetilde{F}(\beta)\|_\infty = \lim_{\beta' \rightarrow \beta} \sup_t \left| \Phi_\tau \left[ \frac{t - x_i^T \beta'}{\widetilde{\sigma}(x_i; \beta')} \right] - \Phi_\tau \left[ \frac{t - x_i^T \beta}{\widetilde{\sigma}(x_i; \beta)} \right] \right| = o_p(1).$$

*Proof* See Čížek and Sadikoglu (2014, Corollary 9).  $\square$

**Lemma 4** Under Assumptions A.1–A.5, A.7, and A.8, it holds for any  $\beta \in B$ ,  $1 \leq s \leq S$ , and  $n \rightarrow \infty$  that

$$E\{\widehat{y}_i^s(\beta) - \widetilde{y}_i^s(\beta)\}^2 = o(1), \quad (37)$$

and for any  $\beta \in B$ ,  $\theta \in \Theta$ , and  $n \rightarrow \infty$ , that

$$E \left| I(\widehat{y}_i^s(\beta) \leq x_i^T \theta) - I(\widetilde{y}_i^s(\beta) \leq x_i^T \theta) \right|^2 = o(1). \quad (38)$$

Furthermore, statements (37) and (38) along with

$$E \left| I(\widehat{y}_i^s(\beta') \leq x_i^T \theta') I(\widehat{y}_i^s(\beta) \leq x_i^T \theta) - I(\widetilde{y}_i^s(\beta') \leq x_i^T \theta') I(\widetilde{y}_i^s(\beta) \leq x_i^T \theta) \right|^2 = o(1), \quad (39)$$

$$E \left| \rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T \theta') I(\widehat{y}_i^s(\beta) \leq x_i^T \theta) - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T \theta') I(\widetilde{y}_i^s(\beta) \leq x_i^T \theta) \right|^2 = o(1), \quad (40)$$

and

$$E \left| \rho_\tau(\widehat{y}_i^s(\beta') \leq x_i^T \theta') \rho_\tau(\widehat{y}_i^s(\beta) \leq x_i^T \theta) - \rho_\tau(\widetilde{y}_i^s(\beta') \leq x_i^T \theta') \rho_\tau(\widetilde{y}_i^s(\beta) \leq x_i^T \theta) \right|^2 = o(1) \quad (41)$$

hold also uniformly with respect to  $\beta, \beta' \in U(\beta^0, \delta)$  and  $\theta, \theta' \in U(\theta^0, \delta)$  for a sufficiently small  $\delta > 0$ .

*Proof* See Čížek and Sadikoglu (2014, Lemma 10).  $\square$

**Lemma 5** Under Assumptions A.1–A.5 and A.6–A.8, it holds for any  $\beta \in B$ ,  $1 \leq s \leq S$ , and  $n \rightarrow \infty$  that

$$|\widehat{\theta}_n^s(\beta) - b(\widetilde{F}(\beta), \beta)| = o_p(1),$$

and in particular,  $|\widehat{\theta}_n^s(\beta^0) - \theta^0| = o_p(1)$ .

*Proof* See Čížek and Sadikoglu (2014, Lemma 11).  $\square$

### Appendix 3: Proofs of the main asymptotic properties

The proofs in this section rely on the notation introduced at the beginning of Appendix 2.

*Proof of Theorem 2* First, we show that  $\beta^0$  is identified. By definition (5), the instrumental criterion yields  $\theta^0$  at the true value of the parameter  $\beta^0$ ,  $b(F_y, \beta^0) = \theta^0$ . Theorem 1 and the construction of  $\tilde{F}$  in (10) then imply  $b(\tilde{F}(\beta^0), \beta^0) = b(F_y, \beta^0) = \theta^0$ . On the other hand, Lemma 1 indicates that, for any  $\beta^1 \neq \beta^0$ ,  $\beta^1 \in B$ , the QR yields different estimates:  $b(\tilde{F}(\beta^1), \beta^1) \neq b(\tilde{F}(\beta^0), \beta^0) = \theta^0$ .

To prove consistency, note that  $\hat{\theta}_n \rightarrow \theta^0 = b(F_y, \beta^0)$  by Theorem 4. Similarly for any  $s = 1, \dots, S$ ,  $\tilde{\theta}_n^s(\beta^0) \rightarrow b(\tilde{F}(\beta^0), \beta^0) = b(F_y, \beta^0) = \theta^0$  and  $\tilde{\theta}_n^s(\beta) \rightarrow b(\tilde{F}(\beta), \beta) \neq \theta^0$  for  $\beta \neq \beta^0$  since, for a given  $\beta \in B$ , the data  $(\tilde{y}_i^s(\beta), x_i)_{i=1}^n$  also satisfy the assumptions of Theorem 4. The same holds also for  $\sum_{s=1}^S \tilde{\theta}_n^s(\beta^0)/S$  as the limits of  $\tilde{\theta}_n^s(\beta)$  are independent of  $s$  and  $S$  is finite. The III criterion (16) is thus a strictly convex function in  $\hat{\theta}_n$  and  $\sum_{s=1}^S \tilde{\theta}_n^s(\beta^0)/S$ , which converges for any  $\beta$  to

$$\arg \min_{\beta \in B} \left[ b(F_y, \beta^0) - b(\tilde{F}(\beta), \beta) \right]^T \Omega \left[ b(F_y, \beta^0) - b(\tilde{F}(\beta), \beta) \right]. \quad (42)$$

Hence by Assumption A.6, any minimizer  $\hat{\beta}_n^{III}$  of (16) satisfies  $\sum_{s=1}^S \tilde{\theta}_n^s(\hat{\beta}_n^{III})/S \rightarrow \theta^0$  in probability as  $n \rightarrow \infty$  (Newey and McFadden 1994, Theorem 2.7). As the link function  $b$  is one-to-one continuous mapping (Assumption A.6) and the parameter space  $B$  is compact (Assumption A.1),  $\hat{\beta}_n^{III}$  has to converge in probability to  $\beta^0$ , which is the unique minimum of (42) (cf. the proof of Theorem 1 in Gouriéroux et al. 1993).

The proof for asymptotic normality of  $\hat{\beta}_n^{III}$  is similar to the proof of Proposition 3 in Gouriéroux et al. (1993), which is however given for a twice continuously differentiable instrumental criterion. By taking the first-order condition of the optimization problem (16) with respect to  $\beta$ , the first-order condition is obtained almost surely:

$$\left[ \frac{1}{S} \sum_{s=1}^S \frac{\partial \tilde{\theta}_n^s(\hat{\beta}_n^{III})^T}{\partial \beta} \right] \Omega \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\hat{\beta}_n^{III}) \right] = 0.$$

Applying the Taylor expansion around  $\beta^0$  to  $\hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\hat{\beta}_n^{III})$ , we obtain analogously to Gouriéroux et al. (1993, Eq. (51)) for some linear combination  $\xi_n$  of  $\beta^0$  and  $\hat{\beta}_n^{III}$  and for  $n \rightarrow \infty$  that

$$\left[ \frac{1}{S} \sum_{s=1}^S \frac{\partial \tilde{\theta}_n^s(\hat{\beta}_n^{III})^T}{\partial \beta} \right] \Omega \left[ \left\{ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0) \right\} - \frac{1}{S} \sum_{s=1}^S \frac{\partial \tilde{\theta}_n^s(\xi_n)^T}{\partial \beta} \{ \hat{\beta}_n^{III} - \beta^0 \} \right] = 0, \quad (43)$$

and due to the full-rank Assumption A.6, that

$$n^{1/2}(\hat{\beta}_n^{III} - \beta^0) = \left( \frac{\partial b^T}{\partial \beta}(\tilde{F}(\beta^0), \beta^0) \Omega \frac{\partial b}{\partial \beta^T}(\tilde{F}(\beta^0), \beta^0) \right)^{-1} \quad (44)$$

$$\times \frac{\partial b^T}{\partial \beta}(\tilde{F}(\beta^0), \beta^0) \Omega n^{1/2} \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0) \right] + o_p(1), \quad (45)$$

where we used  $\partial \tilde{\theta}_n^s(\beta)/\partial \beta^T \rightarrow \partial b(\tilde{F}(\beta), \beta)/\partial \beta^T$  and  $\partial b(\tilde{F}(\hat{\beta}_n^{III}), \hat{\beta}_n^{III})/\partial \beta^T \rightarrow \partial b(\tilde{F}(\beta^0), \beta^0)/\partial \beta^T = D$  as  $n \rightarrow \infty$  in probability (the required continuity of the derivative of the link function and the full rank of its derivative follow from Assumption A.6, whereas the continuity of  $\tilde{F}$  in  $\beta$  follows from its Corollary 1).

Next, by Theorem 4, we have

$$n^{1/2}(\hat{\theta}_n - \theta^0) = -J^{-1} G_n[\varphi_\tau(y_i - x_i^T \theta^0)x_i] + o_p(1). \quad (46)$$

Since  $\{\tilde{y}_i^s(\beta^0), x_i\}_{i=1}^n$  and  $\tilde{\varepsilon}_i^s \sim N(\mu_\tau \cdot \tilde{\sigma}(x_i; \beta), \tilde{\sigma}(x_i; \beta))$  satisfy Assumptions A.1–A.5 and A.7, Theorem 4 also implies

$$n^{1/2}(\tilde{\theta}_n^s(\beta^0) - \theta^0) = -\tilde{J}^{-1} G_n[\varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0)x_i] + o_p(1) \quad (47)$$

for any simulated path  $s = 1, \dots, S$  and  $n \rightarrow \infty$ . Thus, combining (46) and (47) yields

$$\begin{aligned} n^{1/2} \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0) \right] &= -J^{-1} G_n \left[ \varphi_\tau(y_i - x_i^T \theta^0)x_i \right] \\ &\quad + \tilde{J}^{-1} \frac{1}{S} \sum_{s=1}^S G_n \left[ \varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0)x_i \right] + o_p(1) \\ &= \left[ -J^{-1}, \frac{\tilde{J}^{-1}}{S}, \dots, \frac{\tilde{J}^{-1}}{S} \right] \\ &\quad \times \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{bmatrix} \varphi_\tau(y_i - x_i^T \theta^0) \\ \varphi_\tau(\tilde{y}_i^1(\beta^0) - x_i^T \theta^0) \\ \vdots \\ \varphi_\tau(\tilde{y}_i^S(\beta^0) - x_i^T \theta^0) \end{bmatrix} \otimes x_i + o_p(1). \end{aligned} \quad (48)$$

The random variables  $\varphi_\tau(y_i - x_i^T \theta^0)x_i$  and  $\varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0)x_i$  have zero means (by the definition of  $\theta^0$ ) and finite variances and covariances (due to Assumption A.3), which are computed below. By the central limit theorem, the random vector in (48) thus converges in distribution to a normally distributed random vector. Using the notation  $\Sigma$  and  $\tilde{\Sigma}$  from Assumption A.5 and denoting  $\tilde{K}_{0s} = \text{cov}\{\varphi_\tau(y_i - x_i^T \theta^0)x_i, \varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0)x_i\}$  and  $\tilde{K}_{rs} = \text{cov}\{\varphi_\tau(\tilde{y}_i^r(\beta^0) - x_i^T \theta^0)x_i, \varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0)x_i\}$ , which are independent of  $r, s = 1, \dots, S$ , it follows that  $n^{1/2}[\hat{\theta}_n - S^{-1} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0)]$  is asymptotically normal with the variance-covariance matrix

$$\begin{bmatrix} -J^{-1}, \frac{\tilde{J}^{-1}}{S}, \dots, \frac{\tilde{J}^{-1}}{S} \end{bmatrix} \begin{bmatrix} \Sigma & \tilde{K}_{0s} & \tilde{K}_{0s} & \cdots & \tilde{K}_{0s} \\ \tilde{K}_{0s} & \tilde{\Sigma} & \tilde{K}_{rs} & \cdots & \tilde{K}_{rs} \\ \tilde{K}_{0s} & \tilde{K}_{rs} & \tilde{\Sigma} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \tilde{K}_{rs} \\ \tilde{K}_{0s} & \tilde{K}_{rs} & \cdots & \tilde{K}_{rs} & \tilde{\Sigma} \end{bmatrix} \begin{bmatrix} -J^{-1} \\ \tilde{J}^{-1}/S \\ \vdots \\ \tilde{J}^{-1}/S \end{bmatrix}.$$

This matrix can be rewritten as

$$\begin{aligned} & \text{var} \left( n^{1/2} \left\{ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0) \right\} \right) \\ &= J^{-1} \Sigma J^{-1} + \frac{1}{S} \tilde{J}^{-1} \tilde{\Sigma} \tilde{J}^{-1} + \left( 1 - \frac{1}{S} \right) \tilde{J}^{-1} \tilde{K}_{rs} \tilde{J}^{-1} - 2J^{-1} \tilde{K}_{0s} \tilde{J}^{-1}, \quad (49) \end{aligned}$$

where

$$\begin{aligned} \tilde{K}_{rs} &= E[\varphi_\tau(\tilde{y}_i^r(\beta^0) - x_i^T \theta^0) x_i \cdot \varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0) x_i] \\ &= E_x[E[\varphi_\tau(\tilde{y}_i^r(\beta^0) - x_i^T \theta^0) | x_i] \cdot E[\varphi_\tau(\tilde{y}_i^s(\beta^0) - x_i^T \theta^0) | x_i] x_i x_i^T] \\ &= E[\tilde{F}_{\tilde{y}(\beta^0)}(x_i^T \theta^0 | x_i) - \tau)(\tilde{F}_{\tilde{y}(\beta^0)}(x_i^T \theta^0 | x_i) - \tau) x_i x_i^T] \end{aligned}$$

due to the independence of simulated errors  $\{\varepsilon_i^r\}_{i=1}^n$  and  $\{\varepsilon_i^s\}_{i=1}^n$ , and consequently, of the simulated paths  $\{\tilde{y}_i^r(\beta^0)\}_{i=1}^n$  and  $\{\tilde{y}_i^s(\beta^0)\}_{i=1}^n$  (conditionally on  $x_i$ ). Similarly, we can obtain  $\tilde{K}_{0s} = \text{cov}\{[F_y(x_i^T \theta^0 | x_i) - \tau] x_i, [(F_{\tilde{y}(\beta^0)}(x_i^T \theta^0 | x_i) - \tau) x_i]\}$ . By definition (10) of  $\tilde{F}_{\tilde{y}(\beta^0)}$ ,  $\tilde{F}_{\tilde{y}(\beta^0)}(x_i^T \theta^0 | x_i) = F_y(x_i^T \theta^0 | x_i)$  and thus  $\tilde{K}_{0s} = \tilde{K}_{rs} = K$ . Combining (44) and (49) then yields the asymptotic distribution of  $n^{1/2}(\hat{\beta}_n^{III} - \beta^0)$ , which is normal with variance

$$\begin{aligned} & (D^T \Omega D)^{-1} D^T \Omega \\ & \times \left\{ J^{-1} \Sigma J^{-1} + \frac{1}{S} \tilde{J}^{-1} \tilde{\Sigma} \tilde{J}^{-1} + \left( 1 - \frac{1}{S} \right) \tilde{J}^{-1} K \tilde{J}^{-1} - 2J^{-1} K \tilde{J}^{-1} \right\} \Omega D (D^T \Omega D)^{-1} \end{aligned}$$

(recall that  $D = \partial b(\tilde{F}(\beta^0), \beta^0) / \partial \beta^T$ ). Since  $D$  and  $\Omega$  are full rank square matrices by Assumption A.6, the final expression follows from  $(D^T \Omega D)^{-1} = D^{-1} \Omega^{-1} (D^T)^{-1}$ .  $\square$

**Proof of Theorem 3** To establish the asymptotic equivalence result, we first need to prove that the feasible estimator  $\hat{\beta}_n^{FII}$  is consistent. The argument is the same as in the case of the infeasible estimator in Theorem 2 (the first two paragraphs of the proof) provided that we establish  $\hat{\theta}_n^s(\beta^0) \rightarrow b(\tilde{F}(\beta^0), \beta^0) = b(F_y, \beta^0) = \theta^0$  and  $\hat{\theta}_n^s(\beta) \rightarrow b(\tilde{F}(\beta), \beta) \neq \theta^0$  for  $\beta \neq \beta^0$ . This however follows from Lemma 5, stating that  $\hat{\theta}_n^s(\beta) - b(\tilde{F}(\beta), \beta) = o_p(1)$  as  $n \rightarrow \infty$  for any  $\beta \in B$  and  $s = 1, \dots, S$ . Theorem 2.7 of Newey and McFadden (1994) and the continuity of the one-to-one link function thus again implies that  $\hat{\beta}_n^{FII} \rightarrow \beta^0$  in probability as  $n \rightarrow \infty$ .

Next, taking the first-order derivative of the optimization problem (20) with respect to  $\beta$ , the first-order condition is obtained:

$$\left[ \frac{1}{S} \sum_{s=1}^S \frac{\partial \hat{\theta}_n^s (\hat{\beta}_n^{FII})^T}{\partial \beta} \right] \Omega \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \hat{\theta}_n^s (\hat{\beta}_n^{FII}) \right] = 0. \quad (50)$$

Applying the Taylor expansion around  $\beta^0$  to  $\hat{\theta}_n - \sum_{s=1}^S \hat{\theta}_n^s (\hat{\beta}_n^{FII})/S$ , we obtain for some linear combination  $\xi_n$  of  $\beta^0$  and  $\hat{\beta}_n^{FII}$  that

$$\left[ \frac{1}{S} \sum_{s=1}^S \frac{\partial \hat{\theta}_n^s (\hat{\beta}_n^{FII})^T}{\partial \beta} \right] \Omega \left[ \left\{ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \hat{\theta}_n^s (\beta^0) \right\} - \frac{1}{S} \sum_{s=1}^S \frac{\partial \hat{\theta}_n^s (\xi_n)^T}{\partial \beta} \{ \hat{\beta}_n^{FII} - \beta^0 \} \right] = 0. \quad (51)$$

We know from Lemma 5 that  $\hat{\theta}_n^s(\beta) \rightarrow b(\tilde{F}(\beta), \beta)$ . As the consistency of  $\hat{\beta}_n^{FII}$  and Corollary 1 imply that  $\tilde{F}(\beta)$  is continuous in  $\beta \in U(\beta^0, \delta)$  and that  $\|\tilde{F}(\hat{\beta}_n^{FII}) - \tilde{F}(\beta^0)\| = o_p(1)$  as  $n \rightarrow \infty$ , the continuity of the link function (Assumption A.6) implies that  $\hat{\theta}_n^s(\hat{\beta}_n^{FII}) \rightarrow b(\tilde{F}(\beta^0), \beta^0) = \theta^0$ ; similarly,  $\partial \hat{\theta}_n^s(\hat{\beta}_n^{FII})/\partial \beta^T \rightarrow \partial b(\tilde{F}(\beta^0), \beta^0)/\partial \beta^T$ , and  $\partial \hat{\theta}_n^s(\xi_n)/\partial \beta^T \rightarrow \partial b(\tilde{F}(\beta^0), \beta^0)/\partial \beta^T$  in probability for  $n \rightarrow \infty$ .

Consequently, we have due to the full-rank Assumption A.6 that

$$\begin{aligned} & n^{1/2}(\hat{\beta}_n^{FII} - \beta^0) \\ &= \left( \frac{\partial b^T}{\partial \beta}(\tilde{F}(\beta^0), \beta^0) \Omega \frac{\partial b}{\partial \beta^T}(\tilde{F}(\beta^0), \beta^0) \right)^{-1} \frac{\partial b^T}{\partial \beta}(\tilde{F}(\beta^0), \beta^0) \Omega \\ & \quad \times n^{1/2} \left[ \hat{\theta}_n - \frac{1}{S} \sum_{s=1}^S \hat{\theta}_n^s(\beta^0) \right] + o_p(1). \end{aligned} \quad (52)$$

Recall that  $D$  denotes  $\partial b(\tilde{F}(\beta^0), \beta^0)/\partial \beta^T$ . Subtracting (44) from (52) yields

$$\begin{aligned} n^{1/2}(\hat{\beta}_n^{LII} - \hat{\beta}_n^{FII}) &= (D^T \Omega D)^{-1} D^T \Omega \cdot n^{1/2} \left[ \frac{1}{S} \sum_{s=1}^S \tilde{\theta}_n^s(\beta^0) - \frac{1}{S} \sum_{s=1}^S \hat{\theta}_n^s(\beta^0) \right] \\ & \quad + o_p(1). \end{aligned} \quad (53)$$

To prove the theorem, we have to show that the difference  $\sum_{s=1}^S \tilde{\theta}_n^s(\beta^0)/S - \sum_{s=1}^S \hat{\theta}_n^s(\beta^0)/S$  is asymptotically negligible in probability. The estimates  $\tilde{\theta}_n^s(\beta^0)$  and  $\hat{\theta}_n^s(\beta^0)$  are obtained for data  $y_i = \max\{x_i^T \beta^0 + \varepsilon_i^s, 0\}$ , where  $\varepsilon_i^s \sim N(\mu_\tau \sigma(x_i), \sigma(x_i))$  and  $\sigma(x_i)$  represent the conditional variance functions  $\tilde{\sigma}(x_i; \beta^0)$  and  $\hat{\sigma}_n(x_i; \beta^0)$ , respectively. Thus,  $\tilde{\varepsilon}_i^s(\beta^0)|x_i$  follows a normal distribution with mean  $\mu_\tau \tilde{\sigma}(x_i; \beta^0)$  and variance  $\tilde{\sigma}(x_i; \beta^0)$  and  $(x_i, \tilde{\varepsilon}_i^s(\beta^0))$  follows their joint distribution  $P_0$ . On the other hand,  $\hat{\varepsilon}_i^s(\beta^0)|x_i$  is characterized a different conditional distribution, which is a

normal distribution with mean  $\mu_\tau \widehat{\sigma}_n(x_i; \beta^0)$  and variance  $\widehat{\sigma}_n(x_i; \beta^0)$ ; the joint distribution of  $(x_i, \varepsilon_i^s(\beta^0))$  is denoted  $P_n$  (it has the same marginal distribution of  $x_i$  as  $P_0$ ).

We have already established that both  $\widehat{\theta}_n^s(\beta^0)$  and  $\widehat{\theta}_n^s(\beta^0)$  are consistent estimators of  $\theta^0$ . As the variance functions at  $\beta^0$  are bounded (see the introduction of Appendix 2), censored data simulated from  $P_0$  and  $P_n$  satisfy assumptions of Theorem 4 if condition (32) is verified. As all terms of  $\{r(y_i, x_i, \theta') - r(y_i, x_i, \theta'')\}^2$  in condition (32) with their expectations varying with  $P_n$  have the form  $C_2(\theta', \theta'', \theta^0) \rho_\tau(y - x^T \theta') \rho_\tau(y - x^T \theta'')$ ,  $C_1(\theta', \theta'', \theta^0) \rho_\tau(y - x^T \theta) I(y \leq x^T \theta^0)$ , or  $C_0(\theta', \theta'', \theta^0) I(y \leq x^T \theta^0) I(y \leq x^T \theta^0)$  for some deterministic functions  $C_0, C_1, C_2$  and some  $\theta', \theta'' \in U(\theta^0, \delta)$ , Lemma 4 implies the validity of condition (32). Consequently, we can write using the asymptotic linearity result of Theorem 4 for any  $s = 1, \dots, S$

$$\begin{aligned} n^{1/2} [\widehat{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)] \\ = -\widetilde{J}^{-1} G_{n, P_0} [\varphi_\tau(\max\{0, x_i^T \beta^0 + \varepsilon_i^s\} - x_i^T \theta^0) x_i] \\ + \widetilde{J}^{-1} G_{n, P_n} [\varphi_\tau(\max\{0, x_i^T \beta^0 + \varepsilon_i^s\} - x_i^T \theta^0) x_i] + o_p(1) \end{aligned}$$

as  $n \rightarrow \infty$ . Let us denote  $g(x_i, \varepsilon_i^s; \beta, \theta) = \varphi_\tau(\max\{0, \varepsilon_i^s + x_i^T \beta\} - x_i^T \theta) x_i$ . Then

$$\begin{aligned} n^{1/2} [\widehat{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)] \\ = \widetilde{J}^{-1} \left\{ G_{n, P_n} [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] - G_{n, P_0} [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] \right\} + o_p(1). \end{aligned}$$

To verify the claim  $n^{1/2} [\widehat{\theta}_n^s(\beta^0) - \widehat{\theta}_n^s(\beta^0)] = o_p(1)$ , we have to prove that  $G_{n, P_n} [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] - G_{n, P_0} [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] = (G_{n, P_n} - G_{n, P_0}) [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] \xrightarrow{P} 0$  as  $n \rightarrow \infty$ . We only have to show that  $(G_{n, P_n} - G_{n, P_0}) [g(x_i, \varepsilon_i^s; \beta^0, \theta^0)] \rightarrow 0$  in distribution since this is equivalent to the convergence to 0 in probability (cf., the proof of Van der Vaart 2000, Lemma 19.24); see also Čížek and Sadikoglu (2014) for details. This result follows from Van der Vaart and Wellner (1996), Theorem 2.8.9 once we verify their assumptions:

1. The class of functions  $\mathcal{G} = \{g(x, \varepsilon; \beta, \theta) : \beta \in U(\beta^0, \delta), \theta \in U(\theta^0, \delta)\}$ ,  $\delta > 0$ , should satisfy the uniform entropy condition. Since  $g(x, \varepsilon; \beta, \theta) = [\tau - I(\max\{0, x^T \beta + \varepsilon\} - x^T \theta \leq 0)] x_i$ , the class  $\mathcal{G}$  forms a VC class by Van der Vaart and Wellner (1996, Lemma 2.6.18) and it thus satisfies the uniform entropy condition by Van der Vaart and Wellner (1996, Theorem 2.6.7).
2. Next,  $P_n \bar{g}$  has to be bounded for an envelope function  $\bar{g}$  of  $\mathcal{G}$  and has to satisfy the Lindenberg condition  $\limsup_{n \rightarrow \infty} P_n \bar{g}^2 \{\bar{g} \geq \epsilon \sqrt{n}\} = 0$  for every  $\epsilon > 0$ . As the functions in  $\mathcal{G}$  are bounded by a constant (see Assumption A.3), this requirement is also satisfied.
3. Finally, it has to hold that  $\sup_{g, g' \in \mathcal{G}} |E_{P_n}(g - g')^2 - E_{P_0}(g - g')^2| \rightarrow 0$  as  $n \rightarrow \infty$ , where  $E_P(g - g')^2 = E_P[I(\max\{0, x_i^T \beta + \varepsilon_i\} - x_i^T \theta \leq 0) - I(\max\{0, x_i^T \beta' + \varepsilon_i\} - x_i^T \theta' \leq 0)]^2 x_i x_i^T$ ,  $(x_i, \varepsilon_i) \sim P$ , and  $(\beta, \theta)$  and  $(\beta', \theta')$  correspond to func-

tions  $g$  and  $g'$ , respectively. This however directly follows from Lemma 4 as  $x_i$  has a compact support by Assumption A.3.

Consequently,  $n^{1/2} [\tilde{\theta}_n^s(\beta^0) - \hat{\theta}_n^s(\beta^0)] = o_p(1)$  for any  $s = 1, \dots, S$  and it follows from Eq. (53) that  $n^{1/2} (\tilde{\beta}_n^{III} - \hat{\beta}_n^{FII}) = (D' \Omega D)^{-1} D' \Omega o_p(1) = o_p(1)$ .  $\square$

## References

- Arabmazar A, Schmidt P (1982) An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica* 50(4):1055–1063
- Angrist J, Chernozhukov V, Fernandez-Val I (2006) Quantile regression under misspecification, with an application to the US wage structure. *Econometrica* 74(2):539–563
- Brännäs K, Laitila T (1989) Heteroskedasticity in the Tobit model. *Stat Pap* 30(1):185–196
- Buchinsky M, Hahn J (1998) An alternative estimator for the quantile regression model. *Econometrica* 66:653–672
- Campbell J, Honore BE (1993) Median unbiasedness of estimators of panel data censored regression models. *Econom Theory* 9(3):499–503
- Chernozhukov V, Hong H (2002) Three-step censored quantile regression and extramarital affairs. *J Am Stat Assoc* 97:872–882
- Čížek P (2006) Least trimmed squares in nonlinear regression under dependence. *J Stat Plan Inference* 136:3967–3988
- Čížek P, Sadikoglu S (2014) Bias-corrected quantile regression estimation of censored regression models. CentER Discussion Paper 60/2014, Tilburg University, The Netherlands
- Fahr R (2004) Loafing or learning? The demand for informal education. *Eur Econ Rev* 49:75–98
- Fitzenberger B (1997a) A guide to censored quantile regressions. In: Maddala GS, Rao CR (eds) *Handbook of Statistics. Robust Inference*, vol 15. North-Holland, Amsterdam, pp 405–437
- Fitzenberger B (1997b) Computational aspects of censored quantile regression. In: Dodge Y (ed) *Proceedings of the third international conference on statistical data analysis based on the L1-norm and related methods*, vol 31. IMS, Hayward, pp 171–186
- Fitzenberger B, Winker P (2007) Improving the computation of censored quantile regression. *Comput Stat Data Anal* 52:88–108
- Gouriéroux C, Monfort A, Renault E (1993) Indirect inference. *J Appl Econ* 8:85–118
- Gouriéroux C, Renault E, Touzi N (2000) Calibration by simulation for small sample bias correction. In: Mariano RS, Schuermann T, Weeks M (eds) *Simulation-based inference in econometrics: methods and applications*. Cambridge University Press, Cambridge, pp 328–358
- Gouriéroux C, Phillips PCB, Yu J (2010) Indirect inference for dynamic panel models. *J Econom* 157:68–77
- Hall P, Racine JS, Li Q (2004) Cross-validation and the estimation of conditional probability densities. *J Am Stat Assoc* 99:1015–1026
- Honore BE (1992) Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica* 60(3):533–565
- Honore BE, Powell JL (1994) Pairwise difference estimators for censored and truncated regression models. *J Econom* 64:241–278
- Honore B, Khan S, Powell JL (2002) Quantile regression under random censoring. *J Econom* 109:67–105
- Horowitz JL (1986) A distribution-free least squares estimator for linear censored regression models. *J Econom* 32:59–84
- Jonsson R (2012) When does Heckman's two-step procedure for censored data work and when does it not? *Stat Pap* 53(1):33–49
- Khan S, Powell JL (2001) Two-step estimation of semiparametric censored regression models. *J Econom* 103:73–110
- Li Q, Racine JS (2008) Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *J Bus Econ Stat* 26:423–434
- Melenberg B, van Soest AA (1996) Parametric and semi-parametric modeling of vacation expenditures. *J Appl Econ* 11(1):59–76
- Moon C-G (1989) A Monte Carlo comparison of semiparametric Tobit estimators. *J Appl Econ* 4:361–382



- Newey WK, McFadden D (1994) Estimation and hypothesis testing in large samples. In: Engle RF, McFadden D (eds) *Handbook of Econometrics*, vol 4. North-Holland, Amsterdam
- Paarsch HJ (1984) A Monte Carlo comparison of estimators for censored regression models. *J Econ* 24:197–213
- Portnoy S (2003) Censored regression quantiles. *J Am Stat Assoc* 98:1001–1012
- Powell JL (1984) Least absolute deviations estimator for the censored regression model. *J Econom* 25:303–325
- Powell JL (1986a) Censored regression quantiles. *J Econom* 32:143–155
- Powell JL (1986b) Symmetrically trimmed least squares estimation of Tobit models. *Econometrica* 54:1435–1460
- Racine JS, Li Q (2004) Nonparametric estimation of regression functions with both categorical and continuous data. *J Econom* 119:99–130
- Tang Y, Wang HJ, He X, Zhu Z (2011) An informative subset-based estimator for censored quantile regression. *Test* 21:1–21
- Van der Vaart AW (2000) *Asymptotic statistics*. Cambridge University Press, Cambridge
- Van der Vaart AW, Wellner J (1996) *Weak convergence and empirical processes: with applications to statistics*. Springer, New York
- White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25
- Xia Y, Tong H, Li WK, Zhu L-X (2002) An adaptive estimation of dimension reduction space. *J R Stat Soc Ser B* 64(3):363–410