Current status linear regression

Groeneboom, Piet; Hendrickx, Kim

DOI
10.1214/17-AOS1589

Publication date
2018

Document Version
Final published version

Published in
Annals of Statistics

The Annals of Statistics
2018, Vol. 46, No. 4, 1415–1444
https://doi.org/10.1214/17-AOS1589

CURRENT STATUS LINEAR REGRESSION

BY PIET GROENEBOOM AND KIM HENDRICKX[1]

Delft University of Technology and Hasselt University

We construct $\sqrt{n}$-consistent and asymptotically normal estimates for the finite dimensional regression parameter in the current status linear regression model, which do not require any smoothing device and are based on maximum likelihood estimates (MLEs) of the infinite dimensional parameter. We also construct estimates, again only based on these MLEs, which are arbitrarily close to efficient estimates, if the generalized Fisher information is finite. This type of efficiency is also derived under minimal conditions for estimates based on smooth nonmonotone plug-in estimates of the distribution function. Algorithms for computing the estimates and for selecting the bandwidth of the smooth estimates with a bootstrap method are provided. The connection with results in the econometric literature is also pointed out.

1. Introduction. Investigating the relationship between a response variable Y and one or more explanatory variables is a key activity in statistics. Often encountered in regression analysis, however, are situations where a part of the data is not completely observed due to some kind of censoring. In this paper, we focus on modeling a linear relationship when the response variable is subject to interval censoring type I, that is, instead of observing the response Y, one only observes whether or not $Y \leq T$ for some random censoring variable T, independent of Y. This type of censoring is often referred to as the current status model and arises naturally, for example, in animal tumorigenicity experiments (see, e.g., [7] and [8]) and in HIV and AIDS studies (see, e.g., [29]). Substantial literature has been devoted to regression models with current status data including the proportional hazard model studied in [17], the accelerated failure time model proposed by [23] and the proportional odds regression model of [24].

The regression model we want to study is the semiparametric linear regression model $Y = \beta_0' X + \varepsilon$, where the error terms are assumed to be independent of T and X with unknown distribution function F_0. This model is closely related to the binary choice model type, studied in econometrics (see, e.g., [2, 4, 19] and [6]), where, however, the censoring variable T is degenerate, that is, $P(T = 0) = 1$, and observations are of the type $(X_i, 1_{\{Y_i \leq 0\}})$. In the latter model, the scale is not

Received June 2016; revised March 2017.

[1]Supported by the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) and by the Research Foundation Flanders (FWO) Grant 11W7315N.

MSC2010 subject classifications. Primary 62G05, 62N01; secondary 62-04.

Key words and phrases. Current status, linear regression, MLE, semiparametric model.

identifiable, which one usually solves by adding a constraint on the parameter space such as setting the length of β or the first coefficient equal to one.

Our model of interest is parametrized by the finite dimensional regression parameter β_0 and the infinite dimensional nuisance parameter F_0 that contains β_0 as one of its arguments. A similar bundled parameter problem was studied by [5], where the authors first provide a framework for the distributional theory of problems with bundled parameters, and next prove their theory for efficient estimation in the linear regression model with right censored data. A spline based estimate of the nuisance parameter is proposed.

Although it is indeed tempting to think that some kind of smoothing is needed, like the splines in [5] or the kernel estimates in the econometric literature for the binary choice model, where even higher order kernels are used (see, e.g., [19]), a maximum rank correlation estimate, which does not use any smoothing has been introduced in [14], and this estimator has been proved to be $\sqrt{n}$-consistent and asymptotically normal in [28]. However, the latter estimate does not attain the efficiency bounds and one wonders whether it is possible to construct simple discrete estimates of this type and achieve the efficiency bounds. It is not clear how the maximum rank correlation estimate in [14] could be used to this end, and we therefore turn to estimators depending on maximum likelihood estimators for the nuisance parameter.

The profile maximum likelihood estimator (MLE) of β_0 was proved to be consistent in [2] but nothing is known about its asymptotic distribution, apart from its consistency and upper bounds for its rate of convergence. It remains an open question whether or not the profile MLE of β_0 is $\sqrt{n}$-consistent. [22] derived an $n^{1/3}$-rate for the profile MLE; we show that without any smoothing it is possible to construct estimates, based on the MLE for the distribution function F for fixed β, that converge at $\sqrt{n}$-rate to the true parameter. We note, however, that the estimator we propose, based on the nonparametric MLE for F for fixed β, is *not* the profile MLE for β_0. The estimator is a kind of hybrid estimator, which is based on the argmax MLE for F for fixed β, but defined as the zero of a nonsmooth score function as a function of β. So we have the remarkable situation that finding the estimate $\hat{\beta}_n$ as the root of a score equation based on the MLEs $\hat{F}_{n,\beta}$, can be proved to give $\sqrt{n}$-consistent estimates of β_0, in contrast with the argmax approach, using profile likelihood, for which we even still do not know whether it is $\sqrt{n}$-consistent. We go somewhat deeper into this matter in the discussion section of this paper.

A general theoretical framework for semiparametric models when the criterion function is not smooth is developed in [1]. The proposed theory is less suited for our score approach since the authors assume existence of a uniform consistent estimator for the infinite dimensional regression parameter with convergence rate not depending on the finite dimensional regression parameter of interest. In the current status linear regression model, we have to estimate β_0 and F_0 simultaneously, as a consequence the convergence rate of the estimator for F_0 depends on the convergence rate of the estimator for β_0, the parameters β_0 and F_0 are bundled and, therefore, we cannot apply their theory.

[22] considers efficient estimation for the current status model with a 1-dimensional regression parameter β via a penalized maximum likelihood estimator under the conditions that F_0 and $u \mapsto E_\beta(X|T - \beta X = u)$ are three times continuously differentiable and that the data only provide information about a part of the distribution function F_0, where F_0 stays away from zero and 1. [21] proposes an estimation equation for β, derived from an inequality on the conditional covariance between X and Δ conditional on $T - \beta' X$, and uses a U-statistics representation, involving summation over many indices. [27] considers an estimator based on a random sieved likelihood, but the expression for the efficient information (based on the generalized Fisher information) in this paper seems to be different from what we and the authors mentioned above obtain for this expression.

Approaches to $\sqrt{n}$-consistent and efficient estimation of the regression parameters in the binary choice model were considered by [19] and [4] among others. For a derivation of the efficient information $\tilde{\ell}^2_{\beta_0, F_0}$, defined by

$$\begin{aligned}\tilde{\ell}_{\beta,F}(t,x,\delta) &= \{\mathbb{E}(X|T - \beta' X = t - \beta' x) - x\} f(t - \beta' x) \\ &\quad \cdot \left\{ \frac{\delta}{F(t - \beta' x)} - \frac{1 - \delta}{1 - F(t - \beta' x)} \right\},\end{aligned} \tag{1.1}$$

where we assume $f(t - \beta' x) > 0$, we refer to [3] for the binary choice model, and next to [18] and [22] for the current status regression model.

As mentioned above, the condition that the support of the density of $T - \beta' X$ is strictly contained in an interval D for all β in the parameter space and that F_0 stays strictly away from 0 and 1 on D is used in [22]. This condition is also used in [18] and [27]. The drawback of the assumption is that we have no information about the whole distribution F_0. It also goes against the usual conditions made for the current status model, where one commonly assumes that the observations provide information over the whole range of the distribution one wants to estimate. We presume that this assumption is made for getting the Donsker properties to work and to avoid truncation devices that can prevent the problems arising if this condition is not made, such as unbounded score functions and ensuing numerical difficulties. Examples of truncation methods can be found in [4] and [19] among others where the authors consider truncation sequences that converge to zero with increasing sample size. We show that it is possible to estimate the finite dimensional regression parameter β_0 at $\sqrt{n}$-rate based on a fixed truncated subsample of the data where the truncation area is determined by the quantiles of the infinite dimensional nuisance parameter estimator.

The paper is organized as follows. The model, its corresponding log likelihood and a truncated version of the log likelihood are introduced in Section 2. In this section, we also discuss the advantages of a score approach over the maximum likelihood characterization. The behavior of the MLE for the distribution function F_0 in case β is not equal to β_0 is studied in Section 3. We first construct in Section 4, based on a score equation, a $\sqrt{n}$-consistent but inefficient estimate of the

regression parameter based on the MLE of F_0 and show how an estimate of the density, based on the MLE, can be used to extend the estimate of the regression parameter to an estimate with an asymptotic variance that is arbitrarily close to the information lower bound.

Next, we give the asymptotic behavior of a plug-in estimator which is obtained by a score equation derived from the truncated log likelihood in case a second-order kernel estimate for the distribution function F_0 is considered. We show that the latter estimator is $\sqrt{n}$-consistent and asymptotically normal with an asymptotic variance that is arbitrarily (determined by the truncation device) close to the information lower bound, just like the estimator based on the MLE we discussed in the preceding paragraph.

The estimation of an intercept term, that originates from the mean of the error distribution, is outlined in Section 5. Section 6 contains details on the computation of the estimates together with the results of our simulation study; a bootstrap method for selecting a bandwidth parameter is also given. A discussion of our results is given in Section 7. The Appendix contains the derivation of the efficient information given in (1.1). The proofs of the results given in this paper are worked out in the Supplementary Material [10].

2. Model description. Let (T_i, X_i, Δ_i), $i = 1, \ldots, n$ be independent and identically distributed observations from $(T, X, \Delta) = (T, X, 1_{\{Y \le T\}})$. We assume that Y is modeled as

$$Y = \beta_0' X + \varepsilon, \tag{2.1}$$

where β_0 is a k-dimensional regression parameter in the parameter space Θ and ε is an unobserved random error, independent of (T, X) with unknown distribution function F_0. We assume that the distribution of (T, X) does not depend on (β_0, F_0), which implies that the relevant part of the log likelihood for estimating (β_0, F_0) is given by

$$\begin{aligned} l_n(\beta, F) &= \sum_{i=1}^{n} \bigl[\Delta_i \log F\bigl(T_i - \beta' X_i\bigr) + (1 - \Delta_i) \log\bigl\{1 - F\bigl(T_i - \beta' X_i\bigr)\bigr\}\bigr] \\ &= \int \bigl[\delta \log F\bigl(t - \beta' x\bigr) + (1 - \delta) \log\bigl\{1 - F\bigl(t - \beta' x\bigr)\bigr\}\bigr] d\mathbb{P}_n(t, x, \delta), \end{aligned} \tag{2.2}$$

where $\mathbb{P}_n$ is the empirical distribution of the (T_i, X_i, Δ_i). We will denote the probability measure of (T, X, Δ) by P_0. We define the truncated log likelihood $l_n^{(\epsilon)}(\beta, F)$ by

$$\begin{aligned} \int_{F(t-\beta' x) \in [\epsilon, 1-\epsilon]} &\bigl[\delta \log F\bigl(t - \beta' x\bigr) \\ &+ (1 - \delta) \log\bigl\{1 - F\bigl(t - \beta' x\bigr)\bigr\}\bigr] d\mathbb{P}_n(t, x, \delta), \end{aligned} \tag{2.3}$$

where $\epsilon \in (0, 1/2)$ is a truncation parameter. Analogously, let

$$\psi_n^{(\epsilon)}(\beta, F) = \int_{F(t-\beta' x)\in[\epsilon,1-\epsilon]} \phi(t, x, \delta)\{\delta - F(t - \beta' x)\}\, d\mathbb{P}_n(t, x, \delta), \tag{2.4}$$

define the truncated score function for some weight function ϕ. In this paper, we consider estimates of β_0, derived by the idea of solving a score equation $\psi_n^{(\epsilon)}(\beta, \hat{F}_\beta) = 0$ where $\hat{F}_\beta$ is an estimate of F for fixed β. A motivation of the score approach is outlined below. We have three reasons for using the score function characterization instead of the argmax approach for the estimation of β_0:

(i) Our simulation experiments indicate that, even if the profile MLE would be $\sqrt{n}$-consistent, its variance is clearly bigger than the other estimates we propose.

(ii) The characterization of $\hat{\beta}_n$ as the solution of a score equation

$$\psi_n(\beta, \hat{F}_{n,\beta}) = 0$$

[see, e.g., (2.4)], where $\hat{F}_{n,\beta}$ is the MLE for fixed β maximizing the log likelihood defined in (2.2) over all $F \in \mathcal{F} = \{F : \mathbb{R} \to [0, 1] : F \text{ is a distribution function}\}$, gives us freedom in choosing the function ψ_n of which we try to find the root $\hat{\beta}_n$. Smoothing techniques can be used but are not necessary to obtain $\sqrt{n}$-convergence of the estimate.

In this paper, we first choose a function ψ_n, which produces a $\sqrt{n}$-consistent and asymptotically normal estimate of β_0, and does not need any smoothing device. Just like the Han maximum correlation estimate, this estimate does not attain the efficiency bound, although the difference between its asymptotic variance and the efficient asymptotic variance is rather small in our experiments. More details are given in Section 6.

Next, we choose a function ψ_n which gives (only depending on our truncation device) an asymptotic variance which is arbitrarily close to the efficient asymptotic variance. In this case, we need an estimate of the density of the error distribution and are forced to use smoothing in the definition of ψ_n. The estimate, although efficient in the sense we use this concept in our paper, is not necessarily better in small samples, though.

(iii) The "canonical" approach to proofs that argmax estimates of β_0 are $\sqrt{n}$-consistent has been provided by [28]. His Theorem 1 says that $\|\hat{\beta}_n - \beta_0\| = O_p(n^{-1/2})$, where $\|\cdot\|$ denotes the Euclidean norm, if $\hat{\beta}_n$ is the maximizer of $\Gamma_n(\beta)$, with population equivalent $\Gamma(\beta)$ and:

(a) there exists a neighborhood N of β_0 and a constant $k > 0$ such that

$$\Gamma(\beta) - \Gamma(\beta_0) \le -k\|\beta - \beta_0\|^2,$$

for $\beta \in N$, and

(b) uniformly over $o_p(1)$ neighborhoods of β_0,

$$\Gamma_n(\beta) - \Gamma_n(\beta_0) = \Gamma(\beta) - \Gamma(\beta_0) + O_p(\|\beta - \beta_0\|/\sqrt{n}) + o_p(\|\beta - \beta_0\|^2) + O_p(n^{-1}).$$

If we try to apply this to the profile MLE $\hat{\beta}_n$, it is not clear that an expansion of this type will hold. We seem to get inevitably an extra term of order $O_p(n^{-2/3})$ in (b), which does not fit into this framework. On the other hand, in the expansion of our score function ψ_n, we get that this function is in first order the sum of a term of the form

$$\psi'(\beta_0)(\beta - \beta_0),$$

where ψ' is the matrix, representing the total derivative of the population equivalent score function ψ, and a term W_n of order $O_p(n^{-1/2})$, which gives

$$\hat{\beta}_n - \beta_0 \sim -\psi'(\beta_0)^{-1} W_n = O_p\bigl(n^{-1/2}\bigr),$$

and here extra terms of order $O_p(n^{-2/3})$ do not hurt. The technical details are elaborated in the proofs of our main result given in the Supplementary Material [10].

Before we formulate our estimates, we first describe in Section 3 the behavior of the MLE $\hat{F}_{n,\beta}$ for fixed β. Throughout the paper, we illustrate our estimates by a simple simulated data example; we consider the model $Y_i = 0.5X_i + \varepsilon_i$, where the X_i and T_i are independent Uniform$(0, 2)$ and where the ε_i are independent random variables with density $f(u) = 384(u - 0.375)(0.625 - u)1_{[0.375,0.625]}(u)$ and independent of the X_i and T_i. Note that the expectation of the random error $\mathbb{E}(\varepsilon) = 0.5$, our linear model contains an intercept, $\mathbb{E}(Y_i | X_i = x_i) = 0.5 + 0.5x_i$.

REMARK 2.1. We chose the present model as a simple example of a model for which the (generalized) Fisher information is finite. This Fisher information easily gets infinite. For example, if F_0 is the uniform distribution on $[0, 1]$ and X and T (independently) also have uniform distributions on $[0, 1]$ and $\beta = 1/2$, the Fisher information for estimating β is given by

$$\int_{u=0}^{1/2} \frac{(x - 1/2)^2}{u(1-u)}\,dx\,du + \int_{u=1/2}^{1} \frac{\{x - (1-u)\}^2}{u(1-u)}\,dx\,du = \infty.$$

We observed in simulations with the uniform distribution that n times the variance of our estimates (using $\epsilon = 0$) steadily decreases with increasing sample size n, suggesting a faster than $\sqrt{n}$-convergence for the estimate in this model. The theoretical framework for estimation of models with infinite Fisher information falls beyond the scope of this paper. So we chose a model where the ratio $f_0(x)^2/[F_0(x)\{1 - F_0(x)\}]$ stays bounded near the boundary of its support by taking a rescaled version of the density $6x(1 - x)1_{[0,1]}(x)$ for f_0. Note that, if the Fisher information is infinite, our theory still makes sense for the truncated version:

$$\int_{F_0(u)\in[\epsilon,1-\epsilon]} \int_{x=0}^{1} \frac{(x - \mathbb{E}\{X | T - X/2 = u\})^2 f_0(u)^2}{F_0(u)\{1 - F_0(u)\}} \cdot f_{X|T-X/2}(x|u)\, f_{T-X/2}(u)\,dx\,du,$$

corresponding to our truncation of the log likelihood and the score function in the sequel. For completeness, we included the derivation of the Fisher information in the Appendix. These calculations provide more insight in the information loss when one moves from a parametric model where F_0 is known to our semiparametric model with unknown F_0.

3. Behavior of the maximum likelihood estimator. For fixed β, the MLE $\hat{F}_{n,\beta}$ of $l_n(\beta, F)$ is a piecewise constant function with jumps at a subset of $\{T_i - \beta' X_i : i = 1, \ldots, n\}$. Once we have fixed the parameter β, the order statistics on which the MLE is based are the order statistics of the values $U_1^{(\beta)} = T_1 - \beta' X_1, \ldots, U_n^{(\beta)} = T_n - \beta' X_n$ and the values of the corresponding $\Delta_i^{(\beta)}$. The MLE can be characterized as the left derivative of the convex minorant of a cumulative sum diagram consisting of the points $(0, 0)$ and

$$\left(i, \sum_{j=1}^{i} \Delta_{(j)}^{(\beta)}\right), \qquad i = 1, \ldots, n, \tag{3.1}$$

where $\Delta_{(i)}^{(\beta)}$ corresponds to the ith order statistic of the $T_i - \beta' X_i$ (see, e.g., Proposition 1.2 in [13] on page 41). We have

$$\mathbb{P}\{\Delta_i^{(\beta)} = 1 \mid U_i^{(\beta)} = u\} = \int F_0\big(u + (\beta - \beta_0)' x\big) f_{X|T-\beta' X}\big(x | T - \beta' X = u\big)\, dx.$$

Hence, defining

$$F_\beta(u) = \int F_0\big(u + (\beta - \beta_0)' x\big) f_{X|T-\beta' X}\big(x | T - \beta' X = u\big)\, dx, \tag{3.2}$$

we can consider the $\Delta_i^{(\beta)}$ as coming from a sample in the ordinary current status model, where the observations are of the form $(U_i^{(\beta)}, \Delta_i^{(\beta)})$, and where the observation times have density $f_{T-\beta' X}$ and where $\Delta_i^{(\beta)} = 1$ with probability $F_\beta(U_i^{(\beta)})$ at observation $U_i^{(\beta)}$.

REMARK 3.1. Assume that T and X are continuous random variables, then we can write

$$\begin{aligned} F_\beta'(u) = &\int f_0\big(u + (\beta - \beta_0)' x\big) f_{X|T-\beta' X}(x|u)\, dx \\ &+ \int F_0\big(u + (\beta - \beta_0)' x\big) \frac{\partial}{\partial u} f_{X|T-\beta' X}(x|u)\, dx. \end{aligned}$$

Integration by parts on the second term yields

$$\begin{aligned} &\int F_0\big(u + (\beta - \beta_0)' x\big) \frac{\partial}{\partial u} f_{X|T-\beta' X}(x|u)\, dx \\ &\quad = -(\beta - \beta_0)' \int f_0\big(u + (\beta - \beta_0)' x\big) \frac{\partial}{\partial u} F_{X|T-\beta' X}(x|u)\, dx. \end{aligned}$$

This implies

$$F'_\beta(u) = \int f_0\bigl(u + (\beta - \beta_0)'x\bigr)\Bigl\{ f_{X|T-\beta'X}(x|u) - (\beta - \beta_0)' \frac{\partial}{\partial u} F_{X|T-\beta'X}(x|u)\Bigr\}\, dx.$$

Assuming that $u \mapsto f_{X|T-\beta'X}(x|u)$ stays away from zero on the support of f_0, this implies by a continuity argument that F_β is monotone increasing on the support of F'_β for β close to β_0.

Also note that we get from the fact that F_0 is a distribution function with compact support

$$\lim_{u\to-\infty} F_\beta(u) = 0 \quad \text{and} \quad \lim_{u\to\infty} F_\beta(u) = 1.$$

So we may assume that F_β is a distribution function for β close to β_0. If X is discrete, a similar argument can be used to show that F_β is a distribution function for β close to β_0 under the assumption that $u \mapsto P(X = x|T - \beta'X = u)$ stays away from zero on the support of f_0.

A picture of the MLE $\hat{F}_{n,\beta}$, based on the values $T_i - \beta X_i$, and the corresponding function F_β for the model used in our simulation experiment, is shown in Figure 1 and compared with F_0. Note that F_β involves both a location shift and a change in shape of F_0.

For fixed β in a neighborhood of β_0, we can now use standard theory for the MLE from current status theory. The following assumptions are used:

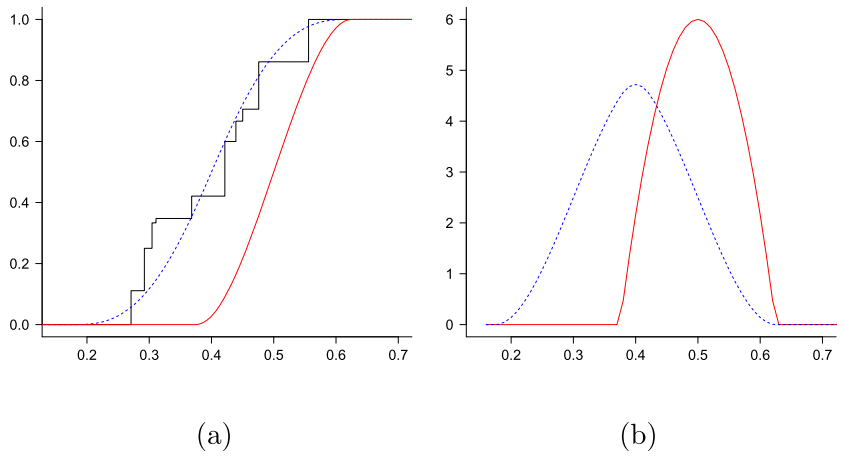

FIG. 1. (a) *The real* F_0 (*red*, *solid*), *the function* F_β *for* $\beta = 0.6$ (*blue*, *dashed*) *and the MLE* $\hat{F}_{n,\beta}$ (*step function*), *for a sample of size* $n = 1000$. (b) *The real* f_0 (*red*, *solid*) *and the function* F'_β *for* $\beta = 0.6$ (*blue*, *dashed*).

(A1) The parameter $\beta_0 = (\beta_{0,1}, \ldots, \beta_{0,k}) \in \mathbb{R}^k$ is an interior point of Θ and the parameter space Θ is a compact convex set.

(A2) F_β has a strictly positive continuous derivative, which stays away from zero on $A_{\epsilon',\beta} \stackrel{\text{def}}{=} \{u : F_\beta(u) \in [\epsilon', 1-\epsilon']\}$ for all $\beta \in \Theta$, where $\epsilon' \in (0, \epsilon)$.

(A3) The density $u \mapsto f_{T-\beta' X}(u)$ is continuous and also staying away from zero on $A_{\epsilon',\beta}$ for all $\beta \in \Theta$, where $A_{\epsilon',\beta}$ is defined as in (A2).

REMARK 3.2. Note that the truncation is for the interval $[\epsilon, 1-\epsilon]$, but that we need conditions (A2) and (A3) to be satisfied for the slightly bigger interval $[\epsilon', 1-\epsilon']$.

LEMMA 3.1. *If Assumptions* (A1), (A2) *and* (A3) *hold, then*:

$$\text{(i)} \qquad \sup_{\beta\in\Theta} \int \{\hat{F}_{n,\beta}(t-\beta' x) - F_\beta(t-\beta' x)\}^2 \, dG(t,x) = O_p(n^{-2/3}).$$

$$\text{(ii)} \qquad \mathbb{P}\Big(\lim_{n\to\infty} \sup_{\beta\in\Theta, u\in A_{\epsilon',\beta}} |\hat{F}_{n,\beta}(u) - F_\beta(u)| = 0\Big) = 1,$$

where ϵ' *is chosen as in condition* (A2).

PROOF. Part (i) is proved in the Supplementary Material. Using the continuity and monotonicity of F_β the second result follows from (i). □

We first show in Section 4.1 that it is possible to construct $\sqrt{n}$-consistent estimates of β_0 derived from a score function $\psi_n^{(\epsilon)}(\beta, \hat{F}_{n,\beta})$ without requiring any smoothing in the estimation process. In Section 4.2, we look at $\sqrt{n}$-consistent and efficient score-estimates based on the MLE $\hat{F}_{n,\beta}$, using a weight function ϕ that incorporates the estimate $\int K_h(u-y)\, d\hat{F}_{n,\beta}(y)$ of the density $f_0(u) = F_0'(u)$. An efficient estimate of β_0 derived by a score function based on kernel estimates for the distribution function, is considered in Section 4.3. The latter estimate does not involve the behavior of the MLE $\hat{F}_{n,\beta}$.

4. $\sqrt{n}$-Consistent estimation of the regression parameter.

4.1. *A simple estimate based on the MLE* $\hat{F}_{n,\beta}$, *avoiding any smoothing.* We consider the function $\psi_{1,n}^{(\epsilon)}$, defined by

$$\psi_{1,n}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} \int_{\hat{F}_{n,\beta}(t-\beta' x)\in[\epsilon,1-\epsilon]} x\{\delta - \hat{F}_{n,\beta}(t-\beta' x)\}\, d\mathbb{P}_n(t,x,\delta), \tag{4.1}$$

where $\hat{F}_{n,\beta}$ is the MLE based on the order statistics of the values $T_i - \beta' X_i$, $i = 1, \ldots, n$. The function $\psi_{1,n}^{(\epsilon)}$ has finitely many different values, and we look for a "crossing of zero" to define our estimate.

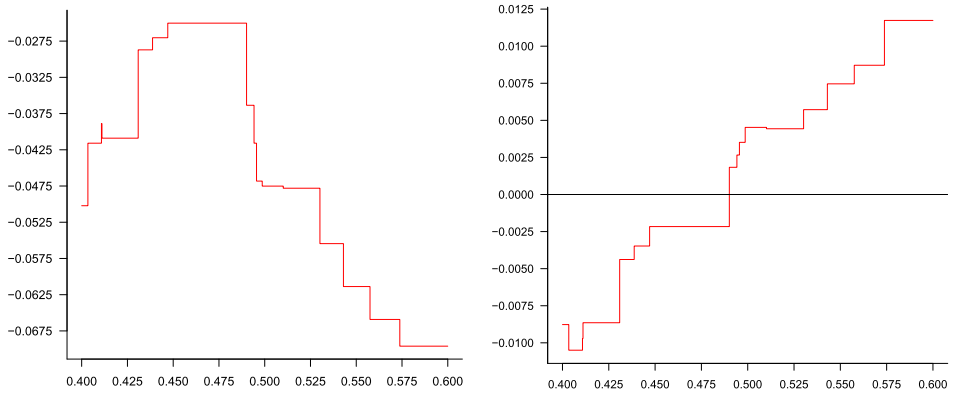

FIG. 2. *The truncated profile log likelihood* $l_n^{(\epsilon)}$ *for the MLE* $\hat{F}_{n,\beta}$ *(left panel) and the score function* $\psi_{1,n}^{(\epsilon)}$ *(right panel) as a function of* β *for a sample of size* $n = 100$ *and* $\epsilon = 0.001$.

Figure 2 gives a picture of the function $\psi_{1,n}^{(\epsilon)}$ as a function of β, drawn as a right-continuous piecewise constant function on a grid of 100 points. Note that this function can have at most $n!$ different values, for all permutations of the numbers $1, \ldots, n$. We would like to define the estimate $\hat{\beta}_n$ by

$$\psi_{1,n}^{(\epsilon)}(\hat{\beta}_n) = 0, \tag{4.2}$$

but it is clear that we cannot hope to achieve that due to the discontinuous nature of the score function $\psi_{1,n}^{(\epsilon)}$. We therefore introduce the following definition.

DEFINITION 4.1 (Zero-crossing). We say that β_* is a crossing of zero of a real-valued function $C : \Theta \to \mathbb{R} : \beta \mapsto C(\beta)$ if each open neighborhood of β_* contains points $\beta_1, \beta_2 \in \Theta$ such that $\bar{C}(\beta_1)\bar{C}(\beta_2) \leq 0$, where $\bar{C}$ is the closure of the image of the function.

We say that a k-dimensional function $\tilde{C} : \Theta \to \mathbb{R}^k : \beta \mapsto \tilde{C}(\beta) = (\tilde{C}_1(\beta), \ldots, \tilde{C}_k(\beta))'$ has a crossing of zero at a point β_* if β_* is a crossing of zero of each component $\tilde{C}_j : \Theta \to \mathbb{R}$, $j = 1 \ldots, k$.

We define our estimator $\hat{\beta}_n$ as a crossing of zero of $\psi_{1,n}^{(\epsilon)}$. Figure 2 shows a crossing of zero at a point β close to $\beta_0 = 0.5$. If the number of dimensions k exceeds one, then a crossing of zero can be thought of as a point $\beta_* \in \Theta$ such that each component of the score function $\psi_{1,n}^{(\epsilon)}$ passes through zero in $\beta = \beta_*$. Before we state the asymptotic result of our estimator in Theorem 4.1, we give in Lemma 4.1 below some interesting properties of the population version of the score function.

LEMMA 4.1. *Let $\psi_{1,\epsilon}$ be defined by*

$$\psi_{1,\epsilon}(\beta) = \int_{F_\beta(t-\beta' x)\in[\epsilon,1-\epsilon]} x\{\delta - F_\beta(t-\beta' x)\}\, dP_0(t,x,\delta), \tag{4.3}$$

and define

$$\mathbb{E}_{\epsilon,\beta}\big(w(T,X,\Delta)\big) = \mathbb{E}\big(1_{\{F_\beta(t-\beta' x)\in[\epsilon,1-\epsilon]\}} w(T,X,\Delta)\big),$$

for functions w, then $\psi_{1,\epsilon}(\beta_0) = 0$ and for each $\beta \in \Theta$ we have:

(i) $\psi_{1,\epsilon}(\beta) = \mathbb{E}_{\epsilon,\beta}\big[\mathrm{Cov}(\Delta, X | T - \beta' X)\big]$,

(ii) $(\beta - \beta_0)' \mathbb{E}_{\epsilon,\beta}\big[\mathrm{Cov}(\Delta, X | T - \beta' X)\big] \geq 0$ *for all* $\beta \in \Theta$,

and β_0 is the only value such that (ii) holds. The derivative of $\psi_{1,\epsilon}$ at $\beta = \beta_0$ is given by

$$\psi'_{1,\epsilon}(\beta_0) = \mathbb{E}_{\epsilon,\beta_0}\big[f_0(T - \beta_0' X)\,\mathrm{Cov}(X | T - \beta_0' X)\big]. \tag{4.4}$$

The proof of Lemma 4.1 is given in the Supplementary Material. An illustration of the second result (ii) is given in Figure 3. This property is used in the proof of consistency of our estimator $\hat{\beta}_n$ given in the Supplementary Material.

The following assumptions are also needed for the asymptotic normality results of our estimators:

(A4) The function F_β is twice continuously differentiable on the interior of the support S_β of $f_\beta = F'_\beta$ for $\beta \in \Theta$.

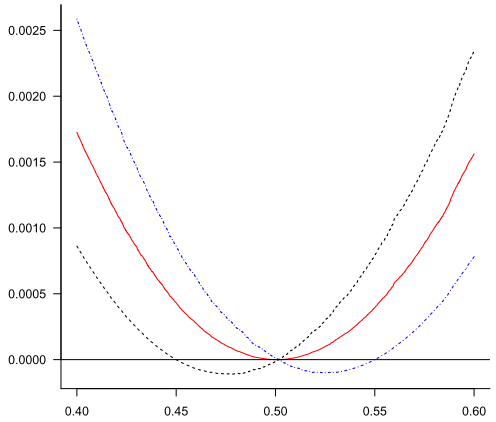

FIG. 3. *The function $\beta \mapsto (\beta - \beta_*)' \int_{F_\beta(t-\beta' x)\in[\epsilon,1-\epsilon]} x\{\delta - F_\beta(t-\beta' x)\}\, dP_0(t,x,\delta)$ as a function of β, with $\beta_* = 0.45$ (black, dashed), $\beta_* = \beta_0 = 0.50$ (red, solid) and $\beta_* = 0.55$ (blue, dashed—dotted) for $\epsilon = 0.001$.*

(A5) The density $f_{T-\beta'X}(u)$ of $T-\beta'X$ and the conditional expectations $\mathbb{E}\{X|T-\beta'X=u\}$ and $\mathbb{E}\{XX'|T-\beta'X=u\}$ are twice continuously differentiable functions w.r.t. u, except possibly at a finite number of points. The functions $\beta\mapsto f_{T-\beta'X}(v)$, $\beta\mapsto\mathbb{E}\{X|T-\beta'X=v\}$ and $\beta\mapsto\mathbb{E}\{XX'|T-\beta'X=v\}$ are continuous functions, for v in the definition domain of the functions and for $\beta\in\Theta$. The density of (T, X) has compact support.

THEOREM 4.1. *Let Assumptions* (A1)–(A5) *be satisfied and suppose that the covariance* $\operatorname{Cov}(X, F_0(u+(\beta-\beta_0)'X)|T-\beta'X=u)$ *is not identically zero for* u *in the region* $A_{\epsilon,\beta}$, *for each* $\beta\in\Theta$. *Moreover, let* $\hat\beta_n$ *be defined by a crossing of zero of* $\psi_{1,n}^{(\epsilon)}$. *Then*:

(i) [Existence of a root] *For all large* n, *a crossing of zero* $\hat\beta_n$ *of* $\psi_{1,n}^{(\epsilon)}$ *exists with probability tending to one.*

(ii) [Consistency]

$$\hat\beta_n\xrightarrow{p}\beta_0,\qquad n\to\infty.$$

(iii) [Asymptotic normality] $\sqrt{n}\{\hat\beta_n-\beta_0\}$ *is asymptotically normal with mean zero and variance* $A^{-1}BA^{-1}$, *where*

$$A=\mathbb{E}_\epsilon\big[f_0(T-\beta_0'X)\operatorname{Cov}\big(X|T-\beta_0'X\big)\big]$$

and

$$B=\mathbb{E}_\epsilon\big[F_0(T-\beta_0'X)\big\{1-F_0(T-\beta_0'X)\big\}\operatorname{Cov}\big(X|T-\beta_0'X\big)\big],$$

defining $\mathbb{E}_\epsilon(w(T,X,\Delta))=\mathbb{E}1_{\{F_0(t-\beta_0'x)\in[\epsilon,1-\epsilon]\}}w(T,X,\Delta)$ *for functions* w *and assuming that* A *is nonsingular.*

REMARK 4.1. Note that $\operatorname{Cov}(X, F_0(u+(\beta-\beta_0)'X)|T-\beta'X=u)$ is not identically zero for u in the region $\{u:\epsilon\le F_\beta(u)\le 1-\epsilon\}$ if the conditional distribution of X, given $T-\beta'X=u$, is nondegenerate for some u in this region if F_0 is strictly increasing on $\{u:\epsilon\le F_\beta(u)\le 1-\epsilon\}$.

The proof of Theorem 4.1 is given in the Supplementary Material [10]. A picture of the truncated profile log likelihood $l_n^{(\epsilon)}(\beta,\hat F_{n,\beta})$ and the score function $\psi_{1,n}^{(\epsilon)}(\beta)$ for β ranging from 0.45 to 0.55 is shown in Figure 2. Note that, since the MLE $\hat F_{n,\beta}$ depends on the ranks of the $T_i-\beta'X_i$, both curves are piecewise constant where jumps are possible if the ordering in $T_i-\beta'X_i$ changes when β changes. Due to the discontinuous nature of the profiled log likelihood and the score function, the estimators are not necessary unique. The result of Theorem 4.1 is valid for any $\hat\beta_n$ satisfying Definition 4.1.

4.2. *Efficient estimates involving the MLE* $\hat{F}_{n,\beta}$. Let K be a probability density function with derivative K' satisfying:

(K1) The probability density K has support $[-1, 1]$, is twice continuously differentiable and symmetric on $\mathbb{R}$.

Let $h > 0$ be a smoothing parameter and K_h respectively K_h' be the scaled versions of K and K', respectively, given by

$$K_h(\cdot) = h^{-1}K\big(h^{-1}(\cdot)\big) \quad \text{and} \quad K_h'(\cdot) = h^{-2}K'\big(h^{-1}(\cdot)\big).$$

The triweight kernel is used in the simulation examples given in the remainder of the paper. Define the density estimate

$$f_{nh,\beta}\big(t - \beta' x\big) = \int K_h\big(t - \beta' x - w\big)\, d\hat{F}_{n,\beta}(w). \tag{4.5}$$

We consider

$$\begin{aligned}\psi_{2,nh}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} &\int_{\hat{F}_{n,\beta}(t-\beta' x)\in[\epsilon,1-\epsilon]} x f_{nh,\beta}\big(t - \beta' x\big) \\ &\cdot \frac{\delta - \hat{F}_{n,\beta}(t - \beta' x)}{\hat{F}_{n,\beta}(t - \beta' x)\{1 - \hat{F}_{n,\beta}(t - \beta' x)\}}\, d\mathbb{P}_n(t, x, \delta),\end{aligned} \tag{4.6}$$

and let, analogously to the first estimator defined in the previous section, $\hat{\beta}_n$ be the estimate of β_0, defined by a zero-crossing of the score function $\psi_{2,nh}^{(\epsilon)}$.

THEOREM 4.2. *Suppose that the conditions of Theorem 4.1 hold and that the function F_β is three times continuously differentiable on the interior of the support S_β. Let $\hat{\beta}_n$ be defined by a zero-crossing of $\psi_{2,nh}^{(\epsilon)}$. Then, as $n \to \infty$, and $h \asymp n^{-1/7}$:*

(i) [Existence of a root] *For all large n, a crossing of zero $\hat{\beta}_n$ of $\psi_{2,nh}^{(\epsilon)}$ exists with probability tending to one.*

(ii) [Consistency]

$$\hat{\beta}_n \xrightarrow{p} \beta_0, \qquad n \to \infty.$$

(iii) [Asymptotic normality] $\sqrt{n}\{\hat{\beta}_n - \beta_0\}$ *is asymptotically normal with mean zero and variance $I_\epsilon(\beta_0)^{-1}$, where*

$$I_\epsilon(\beta_0) = \mathbb{E}_\epsilon\left\{\frac{f_0(T - \beta_0' X)^2 \operatorname{Cov}(X|T - \beta_0' X)}{F_0(T - \beta_0' X)\{1 - F_0(T - \beta_0' X)\}}\right\}, \tag{4.7}$$

which is assumed to be nonsingular.

A picture of the score function $\psi_{2,nh}^{(\epsilon)}(\beta)$ is shown in Figure 4. Note that the range on the vertical axis is considerably larger than the range on the vertical axis of the corresponding score function $\psi_{1,n}^{(\epsilon)}(\beta)$.

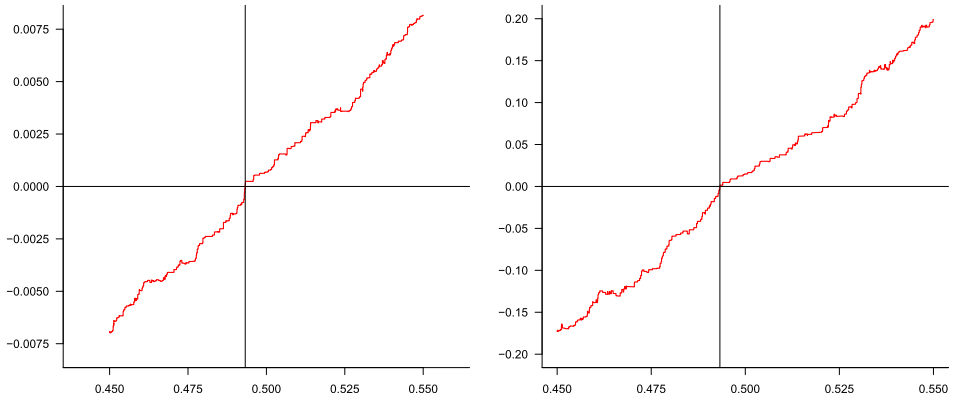

FIG. 4. *The score functions* $\psi_{1,n}^{(\epsilon)}$ *(left panel) and* $\psi_{2,nh}^{(\epsilon)}$ *(right panel) as a function of* β *for a sample of size* $n = 1000$ *with* $\epsilon = 0.001$ *and* $h = 0.5n^{-1/7}$.

4.3. *Efficient estimates not involving the MLE* $\hat{F}_{n,\beta}$. Define the plug-in estimate

$$F_{nh,\beta}(t-\beta' x) = \frac{\int \delta K_h(t-\beta' x-u+\beta' y)\, d\mathbb{P}_n(u,y,\delta)}{\int K_h(t-\beta' x-u+\beta' y)\, d\mathbb{G}_n(u,y)}, \tag{4.8}$$

where $\mathbb{G}_n$ is the empirical distribution function of the pairs (T_i, X_i) and where K_h is a scaled version of a probability density function K, satisfying condition (K1); the probability measure of (T, X) will be denoted by G. The plug-in estimates are not necessarily monotone but we show in Theorem 4.4 that $F_{nh,\beta}$ is monotone with probability tending to one as $n \to \infty$ and $\beta \to \beta_0$. Another way of writing $F_{nh,\beta}$ is in terms of ordinary sums. Let

$$g_{nh,1,\beta}(t-\beta' x) = \frac{1}{n}\sum_{j=1}^{n} \Delta_j K_h(t-\beta' x - T_j + \beta' X_j) \tag{4.9}$$

and

$$g_{nh,\beta}(t-\beta' x) = \frac{1}{n}\sum_{j=1}^{n} K_h(t-\beta' x - T_j + \beta' X_j), \tag{4.10}$$

then

$$F_{nh,\beta}(t-\beta' x) = \frac{g_{nh,1,\beta}(t-\beta' x)}{g_{nh,\beta}(t-\beta' x)} = \frac{\sum_{j=1}^{n} \Delta_j K_h(t-\beta' x - T_j + \beta' X_j)}{\sum_{j=1}^{n} K_h(t-\beta' x - T_j + \beta' X_j)},$$

in which we recognize the Nadaraya–Watson statistic. One could also omit the diagonal term $j = i$ in the sums above when estimating $F_{nh,\beta}(T_i - \beta' X_i)$, which is often done in the econometric literature (see, e.g., [15]). In our computer experiments, however, this gave an estimate of the distribution function, which had a more irregular behavior than the estimator with the diagonal term included.

If we replace F in (2.3) by $F_{nh,\beta}$, the truncated log likelihood becomes a function of β only. Although the log likelihood has discontinuities if we consider the lower and upper boundaries $F_{nh,\beta}^{-1}(\epsilon)$ and $F_{nh,\beta}^{-1}(1-\epsilon)$ of the integral also as a function of β, an asymptotic representation of the partial derivatives of the truncated log likelihood is given by the score function

$$\psi_{3,nh}^{(\epsilon)}(\beta) \stackrel{\text{def}}{=} \int_{F_{nh,\beta}(t-\beta' x)\in[\epsilon,1-\epsilon]} \partial_\beta F_{nh,\beta}(t-\beta' x) \cdot \frac{\delta - F_{nh,\beta}(t-\beta' x)}{F_{nh,\beta}(t-\beta' x)\{1-F_{nh,\beta}(t-\beta' x)\}}\, d\mathbb{P}_n(t,x,\delta), \tag{4.11}$$

where the partial derivative of the plug-in estimate $F_{nh,\beta}(t-\beta' x)$, given by (4.8), w.r.t. β has the following form:

$$\partial_\beta F_{nh,\beta}(t-\beta' x) = \frac{\int (y-x)\{\delta - F_{nh,\beta}(t-\beta' x)\}K_h'(t-\beta' x - u + \beta' y)\, d\mathbb{P}_n(u,y,\delta)}{g_{nh,\beta}(t-\beta' x)}, \tag{4.12}$$

where $g_{nh,\beta}(t-\beta' x)$ is defined in (4.10). We define the plug-in estimator $\hat{\beta}_n$ of β_0 by

$$\psi_{3,nh}^{(\epsilon)}(\hat{\beta}_n) = 0. \tag{4.13}$$

A picture of the truncated log likelihood $l_n^{(\epsilon)}(\beta, F_{nh,\beta})$ and score function $\psi_{3,nh}^{(\epsilon)}(\beta)$ for the plug-in method is shown in Figure 5. Since $F_{nh,\beta}(t-\beta' x)$ is continuous, we no longer need to introduce the concept of a zero-crossing to ensure existence

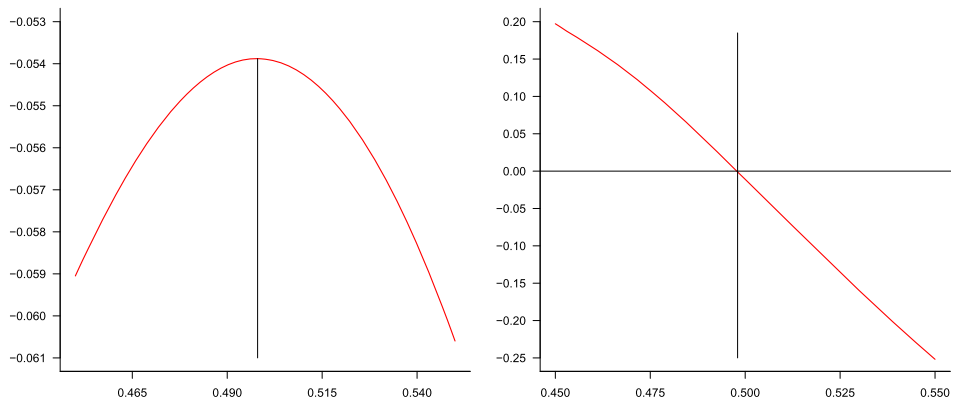

FIG. 5. *The truncated profile log likelihood* $l_n^{(\epsilon)}$ *for the plug-in* $F_{nh,\beta}$ (*left panel*) *and the score function* $\psi_{3,nh}^{(\epsilon)}$ (*right panel*) *as a function of* β *for a sample of size* $n = 1000$ *with* $\epsilon = 0.001$ *and* $h = 0.5n^{-1/5}$.

of the estimator and we can work with the zero of the score function $\psi_{3,nh}^{(\epsilon)}(\beta)$. Our main result on the plug-in estimator is given below.

THEOREM 4.3. *If Assumptions* (A1)–(A5) *hold and*

$$
\begin{aligned}
&-(\beta-\beta_0)'\int_{F_\beta(t-\beta'x)\in[\epsilon,1-\epsilon]}\partial_\beta F_\beta\bigl(t-\beta'x\bigr)\\
&\qquad\cdot\frac{F_0(t-\beta_0'x)-F_\beta(t-\beta'x)}{F_\beta(t-\beta'x)\{1-F_\beta(t-\beta'x)\}}\,dG(t,x),
\end{aligned}
\tag{4.14}
$$

is nonzero for each $\beta\in\Theta$ except for $\beta=\beta_0$, then for $\hat\beta_n$ being the plug-in estimator introduced above, as $n\to\infty$, and $h\asymp n^{-1/5}$:

(i) [Existence of a root] *For al large n a point $\hat\beta_n$, satisfying* (4.13), *exists with probability tending to one.*

(ii) [Consistency]

$$\hat\beta_n\xrightarrow{p}\beta_0,\qquad n\to\infty.$$

(iii) [Asymptotic normality] $\sqrt{n}\{\hat\beta_n-\beta_0\}$ *is asymptotically normal with mean zero and variance $I_\epsilon(\beta_0)^{-1}$ where $I_\epsilon(\beta_0)$, defined in* (4.7), *is assumed to be nonsingular.*

REMARK 4.2. Note that using an expansion in $\beta-\beta_0$, we can write $\partial_\beta F_\beta(t-\beta'x)$ as

$$
\begin{aligned}
&\int(y-x)\,f_0\bigl(t-\beta_0'x+(\beta-\beta_0)'(y-x)\bigr)\,f_{X|T-\beta'X}\bigl(y|T-\beta'X=t-\beta'x\bigr)\,dy\\
&\qquad+\int F_0\bigl(t-\beta_0'x+(\beta-\beta_0)'(y-x)\bigr)\\
&\qquad\cdot\partial_\beta f_{X|T-\beta'X}\bigl(y|T-\beta'X=t-\beta'x\bigr)\,dy\\
&\quad=f_0\bigl(t-\beta'x\bigr)\mathbb{E}\bigl\{X-x|T-\beta'X=t-\beta'x\bigr\}+O(\beta-\beta_0)
\end{aligned}
$$

so that the integral defined in (4.14) can be approximated by

$$
\begin{aligned}
&-(\beta-\beta_0)'\int_{F_\beta(u)\in[\epsilon,1-\epsilon]}f_0(u)\mathbb{E}\bigl\{X-x|T-\beta'X=u\bigr\}\\
&\qquad\cdot\frac{F_0(u+(\beta-\beta_0)'x)-F_\beta(u)}{F_\beta(u)\{1-F_\beta(u)\}}\,f_{X|T-\beta'X}(x|u)\,dx\,du\\
&\quad=\int_{F_\beta(u)\in[\epsilon,1-\epsilon]}\bigl(f_0(u)\operatorname{Cov}\bigl((\beta-\beta_0)'X,F_0\bigl(u+(\beta-\beta_0)'X\bigr)|T-\beta'X=u\bigr)\bigr)\\
&\qquad\cdot\bigl(F_\beta(u)\bigl\{1-F_\beta(u)\bigr\}\bigr)^{-1}du,
\end{aligned}
$$

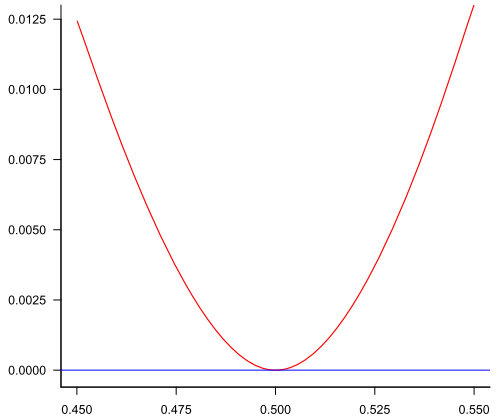

FIG. 6. *The integral defined in* (4.14), *as a function of* β, *with* $\epsilon = 0.001$.

which is positive by the monotonicity of F_0. (See also the discussion in [21] about this covariance and the proof of Lemma 4.1 given in the Supplementary Material.) A crucial property of the covariance used here, showing that the covariance is nonnegative, goes back to a representation of the covariance in [16], which can easily be proved by an application of Fubini's theorem:

$$EXY - EXEY = \int \{\mathbb{P}(X \geq s, Y \geq t) - \mathbb{P}(X \geq s)\mathbb{P}(Y \geq t)\}\, ds\, dt,$$

if XY, X and Y are integrable. Figure 6 shows the integral in (4.14) for our simulation model for $\beta \in [0.45, 0.55]$ and illustrates that this integral is strictly positive except for $\beta = \beta_0$, which is a crucial property for the proof of the consistency of the plug-in estimator given in the Supplementary Material [10].

Section 4.3.1 contains a road map of the proof of Theorem 4.3, the proof itself is given in the Supplementary Material [10]. We also have the following results for the plug-in estimate.

THEOREM 4.4. *Let the conditions of Theorem* 4.3 *be satisfied, then we have on each interval* I *contained in the support of* f_β *and for each* $\beta \in \Theta$

$$P\{F_{nh,\beta} \text{ is monotonically increasing on } I\} \xrightarrow{p} 1.$$

The proof of Theorem 4.4 follows from the asymptotic monotonicity of the plug-in estimate in the classical current status model (without regression parameters) and is proved in the same way as Theorem 3.3 of [12].

THEOREM 4.5. *Let the conditions of Theorem 4.3 be satisfied. Then, for* $\hat{\beta}_n$ *being the plug-in estimator of* β_0,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{I_\epsilon(\beta_0)^{-1}}{\sqrt{n}} \sum_{i \in J_{F_0}} f_0(T_i - \beta_0' X_i)\{\mathbb{E}(X_i | T_i - \beta_0' X_i) - X_i\} \cdot \frac{\Delta_i - F_0(T_i - \beta_0' X_i)}{F_0(T_i - \beta_0' X_i)\{1 - F_0(T_i - \beta_0' X_i)\}} + o_p(1),$$

where $J_H = \{i : \epsilon \le H(T_i - \beta_0' X_i) \le 1 - \epsilon\}$ *for some function* H.

The representation of Theorem 4.5 plays an important role in determining the variance of smooth functionals, of which the intercept $\alpha = \int u \, dF_0(u)$ is an example. The proof of Theorem 4.5 is given in the Supplementary Material [10]. A similar representation holds for the estimators defined in Theorem 4.1 and Theorem 4.2 (see the proofs of Theorem 4.1 and 4.2, respectively, given in the Supplementary Material).

REMARK 4.3. The plug-in method also suggests the use of U-statistics. By straightforward calculations, we can write the score function defined in (4.11) as

$$\begin{aligned}
&\psi_{3,nh}^{(\epsilon)}(\beta) \\
&\quad = \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \frac{\frac{\partial}{\partial \beta} F_{nh,\beta}(T_i - \beta' X_i)\{\Delta_i - F_{nh,\beta}(T_i - \beta' X_i)\}}{F_{nh,\beta}(T_i - \beta' X_i)\{1 - F_{nh,\beta}(T_i - \beta' X_i)\}} \\
&\quad = \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \ne i} \frac{\Delta_i \Delta_j (X_j - X_i) K_h'(T_i - \beta' X_i - T_j + \beta' X_j)}{g_{nh,1,\beta}(T_i - \beta' X_i)} \\
&\qquad + \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \ne i} \big((1 - \Delta_i)(1 - \Delta_j)(X_j - X_i) \\
&\qquad \cdot K_h'\big(T_i - \beta' X_i - T_j + \beta' X_j\big)\big)\big(g_{nh,0,\beta}\big(T_i - \beta' X_i\big)\big)^{-1} \\
&\qquad - \frac{1}{n^2} \sum_{i \in J_{F_{nh,\beta}}} \sum_{j \ne i} \frac{(X_j - X_i) K_h'(T_i - \beta' X_i - T_j + \beta' X_j)}{g_{nh,\beta}(T_i - \beta' X_i)},
\end{aligned} \tag{4.15}$$

where $g_{nh,0,\beta} = g_{nh,\beta} - g_{nh,1,\beta}$; see (4.9) and (4.10). Each of the three terms on the right-hand side of (4.15) can be rewritten in terms of a scaled second order U-statistics. A proof based on U-statistics requires lengthy and tedious calculations, which are avoided in the current approach for proving Theorem 4.3. The representation given in Theorem 4.5 also indicates that the U-statistics representation does not give the most natural approach to the proof of asymptotic normality

and efficiency of $\hat{\beta}_n$. For these reasons, we do not further examine the results on U-statistics.

REMARK 4.4. We propose the bandwidths $h \asymp n^{-1/7}$ respectively $h \asymp n^{-1/5}$ in Theorem 4.2, respectively Theorem 4.3, which are the usual bandwidths with ordinary second-order kernels for the estimates of a density, respectively distribution function. Unfortunately, various advices are given in the literature on what smoothing parameters one should use. [19] has fourth-order kernels and uses bandwidths between the orders $n^{-1/6}$ and $n^{-1/8}$ for the estimation of F. Note that the use of fourth-order kernels needs the associated functions to have four derivatives in order to have the desired bias reduction. [4] advises a bandwidth h such that $n^{-1/5} \ll h \ll n^{-1/8}$, excluding the choice $h \asymp n^{-1/5}$. Both ranges are considerably large and exclude our bandwidth choice $h \asymp n^{-1/5}$. [22] considers a penalized maximum likelihood estimator where the penalty parameter λ_n satisfies $1/\lambda_n = O_p(n^{2/5})$ and $\lambda_n^2 = o_p(n^{-1/2})$. Translated into bandwidth choice (using $h_n \asymp \sqrt{\lambda_n}$), the conditions correspond to: $n^{-1/5} \lesssim h \ll n^{-1/8}$, suggesting that their conditions do allow the choice $h \asymp n^{-1/5}$ for estimating the distribution function.

4.3.1. *Road map of the proof of Theorem* 4.3. The older proofs of a result of this type always used second derivative calculations. As convincingly argued in [30], proofs of this type should only use first derivatives and that is indeed what we do. The limit function F_β of the estimates for F_0 when $\beta \neq \beta_0$ plays a crucial role in our proofs. We first prove the consistency of the plug-in estimate $\hat{\beta}_n$. Next, we use a Donsker property for the functions representing the score function and prove that the integral w.r.t. $d\mathbb{P}_n$ of this score function is

$$o_p(n^{-1/2} + \hat{\beta}_n - \beta_0),$$

and that the integral w.r.t. dP_0 is asymptotically equivalent to

$$-(\hat{\beta}_n - \beta_0) I_\epsilon(\beta_0),$$

where $I_\epsilon(\beta_0)$ is the generalized Fisher information, given by (4.7). Combining these results give Theorem 4.3. Very essential in this proof are L_2-bounds on the distance between the functions $F_{nh,\beta}$ to its limit F_β for fixed β and on the L_2-distance between the first partial derivatives $\partial_\beta F_{nh,\beta}$ and $\partial_\beta F_\beta$. If the bandwidth $h \asymp n^{-1/5}$, the first L_2-distance is of order $n^{-2/5}$ and the second distance is of order $n^{-1/5}$, allowing us to use the Cauchy–Schwarz inequality on these components. Here, we use a result in [9] on L_2 bounds of derivatives of kernel density estimates.

In Section 5, we discuss the estimation of an efficient estimate of the intercept term in regression model (2.1) using the plug-in estimates $\hat{\beta}_n$ and $F_{nh,\hat{\beta}_n}$.

4.4. *Truncation*. We introduced a truncation device in order to avoid unbounded score functions and numerical difficulties. If one starts with the *efficient* score equation or an estimate thereof, the solution sometimes suggested in the literature, is to add a constant c_n, tending to zero as $n \to \infty$, to the factor $F(t - \beta' x)\{1 - F(t - \beta' x)\}$, which inevitably will appear in the denominator. This is done in, for example, [21]; similar ideas involving a sequence (c_n) are used in [19] and [4].

In contrast with the usual approaches to truncation, which imply the selection of a suitable sequence c_n, we do not consider a vanishing truncation sequence but work with a subsample of the data depending on the ϵ and $(1 - \epsilon)$ quantiles of the distribution function estimate for small but fixed $\epsilon \in (0, 1/2)$. This simple device in (2.3) moreover implies keeping the characterizing properties of the MLE (see Proposition 1.1 on page 39 of [13]), which are lost when a vanishing sequence is considered. It is perhaps somewhat remarkable that we can, instead of letting $\epsilon \downarrow 0$, fix $\epsilon > 0$ and still have consistency of our estimators; on the other hand, the estimate proposed by [22] is also identified via a subset of the support of the distribution F_0.

Although the truncation area depends on β, we show in the Supplementary Material [10] (see the proof of Theorem 4.1) that the population version of the score function, given by

$$\psi_\epsilon(\beta) = \int_{F_\beta(t-\beta' x)\in[\epsilon,1-\epsilon]} \phi(t, x, \delta)\{\delta - F_\beta(t - \beta' x)\}\, dP_0(t, x, \delta), \tag{4.16}$$

has a derivative at $\beta = \beta_0$ that only involves the derivative of the integrand in (4.16), but does not involve terms arising from the truncation limits appearing in the integral. Using the truncation in the argmax maximum log likelihood approach would not lead to a derivative of the population version of the log likelihood, which ignores the boundaries and, therefore, this truncation is less suited for argmax estimators.

A drawback of our fixed truncation parameter approach is that we get a truncated Fisher information. The resulting estimates are therefore not efficient in the classical sense of efficiency but the difference between the efficient variance and almost (determined by the size of ϵ) efficient variance is rather small in our simulation models. We also tried to program the fully efficient estimators proposed by [21] and compared its performance to the performance of our almost-efficient estimators. The comparison showed that our estimates perform better in finite samples. Moreover, the estimates by [21] involve several kernel density estimates, resulting in a very large computation time compared to our simple estimates (involving 5 double summations over the data points).

Moreover, the usual conditions in the theory of estimation of F_0 under current status and, more generally, interval censored data are that F_0 corresponds to a distribution with compact support. Otherwise, certain variances easily get infinite, and

similarly, the Fisher information in our model can easily become infinite. Truncating by keeping the quantiles between ϵ and $1-\epsilon$ avoids difficulties in this case and allows us to apply the theory, which presently has been developed for the current status model.

Note that the score function defined in (4.1) does not contain a factor $F(t-\beta'x)$ or $1-F(t-\beta'x)$ in the denominator. For simplicity of the proofs, we still impose the truncation area, since the classical results for the current status model are derived under the assumption that the density f_0 is bounded away from zero. We conjecture however that the result of Theorem 4.1 remains valid when taking $\epsilon = 0$.

5. Estimation of the intercept. We want to estimate the intercept

$$\alpha = \int u \, dF_0(u). \tag{5.1}$$

We can take the plug-in estimate $\hat{\beta}_n$ of β_0, by using a bandwidth of order $n^{-1/5}$ and the score procedure, as before. However, in estimating α, as defined by (5.1), we have to estimate F_0 with a smaller bandwidth h, satisfying $h \ll n^{-1/4}$ to avoid bias, for example, $h \asymp n^{-1/3}$. The matter is discussed in [4], page 1253.

We have the following result of which the proof can be found in the Supplementary Material [10].

THEOREM 5.1. *Let the conditions of Theorem 4.3 be satisfied, and let $\hat{\beta}_n$ be the k-dimensional estimate of β_0 as obtained by the score procedure, described in Theorem 4.3, using a bandwidth of order $n^{-1/5}$. Let $F_{nh,\hat{\beta}_n}$ be a plug-in estimate of F_0, using $\hat{\beta}_n$ as the estimate of β_0, but using a bandwidth h of order $n^{-1/3}$ instead of $n^{-1/5}$. Finally, let $\hat{\alpha}_n$ be the estimate of α, defined by*

$$\int u \, dF_{nh,\hat{\beta}_n}(u).$$

Then $\sqrt{n}(\hat{\alpha}_n - \alpha)$ is asymptotically normal, with expectation zero and variance

$$\sigma^2 \stackrel{\text{def}}{=} a(\beta_0)' I_\epsilon(\beta_0)^{-1} a(\beta_0) + \int \frac{F_0(v)\{1-F_0(v)\}}{f_{T-\beta_0'X}(v)} \, dv, \tag{5.2}$$

where $a(\beta_0)$ is the k-dimensional vector, defined by

$$a(\beta_0) = \int \mathbb{E}\{X | T - \beta_0'X = u\} f_0(u) \, du,$$

and $I_\epsilon(\beta_0)$ is defined in (4.7).

REMARK 5.1. We choose the bandwidth of order $n^{-1/3}$ for specificity, but other choices are also possible. We can in fact choose $n^{-1/2} \ll h \ll n^{-1/4}$. The bandwidth of order $n^{-1/3}$ corresponds to the automatic bandwidth choice of the MLE of F_0, also using the estimate $\hat{\beta}_n$ of β_0.

REMARK 5.2. Note that the variance corresponds to the information lower bound for smooth functionals in the binary choice model, given in [4]. The second part of the expression for the variance on the right-hand side of (5.2) is familiar from current status theory; see, for example, (10.7), page 287 of [11].

Instead of considering the plug-in estimate, we could also consider the estimates described in Theorem 4.1 and Theorem 4.2. After having determined an estimate $\hat{\beta}_n$ in this way, we next estimate α by

$$\hat{\alpha}_n = \int x \, d\hat{F}_{n,\hat{\beta}_n}(x), \tag{5.3}$$

where $\hat{F}_{n,\hat{\beta}_n}$ is the MLE corresponding to the estimate $\hat{\beta}_n$. The theoretical justification of this approach can be proved using the asymptotic theory of smooth functionals given in [11], page 286. Using the MLE $\hat{F}_{n,\hat{\beta}_n}$ instead of the plug-in $F_{nh,\hat{\beta}_n}$ as an estimate of the distribution function F_0, avoids the selection of a bandwidth parameter for the intercept estimate. We discuss in the next section how the bandwidth can be selected by the practitioner in a real data sample.

6. Computation and simulation. The computation of our estimates is relatively straightforward in all cases. For the score-estimates defined in Sections 4.1 and 4.2, we first compute the MLE for fixed β by the so-called "pool adjacent violators" algorithm for computing the convex minorant of the "cusum diagram" defined in (3.1). If the MLE has been computed for fixed β, we can compute the density estimate f_{nh}. The estimate of β_0 is then determined by a root-finding algorithm such as Brent's method. Computation is very fast. For the plug-in estimate, we simply compute the estimate $F_{nh,\beta}$ as a ratio of two kernel estimators for fixed β and then compute the derivative w.r.t. β. Next, we use again a root-finding algorithm to determine the zero of the corresponding score function.

Some results from the simulations of our model are available in Table 1, which contains the mean value of the estimate, averaged over $N = 10{,}000$ iterations, and n times the variance of the estimate of $\beta_0 = 0.5$ (resp., $\alpha_0 = 0.5$) for the different methods described above, as well as for the classical MLE of β_0, for different sample sizes n and a truncation parameter $\epsilon = 0.001$. We took the bandwidth $h = 0.5n^{-1/7}$ for the efficient score-estimate of Section 4.2. The bandwidth $h = 0.5n^{-1/5}$ for the plug-in estimate of Section 4.3 was chosen based on an investigation of the mean squared error (MSE) for different choices of c in $h = cn^{-1/5}$. Details on how to choose the bandwidth in practice are given in Section 6.1. The true asymptotic values for the variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ in our simulation model, obtained via the inverse of the Fisher information $I_\epsilon(\beta_0)$, are 0.151707 without truncation and 0.158699 for $\epsilon = 0.001$ and 0.17596 for $\epsilon = 0.01$. We advise to use a truncation parameter ϵ of 0.001 or smaller in practice. The variance defined in Theorem 4.1 for $\epsilon = 0.001$ is 0.193612. The lower bounds for the variance of the

TABLE 1

The mean value of the estimate and n times the variance of the estimates of β_0 and α_0 for different methods, $h_\beta = 0.5n^{-1/7}$ (for the efficient score method), $h_\beta = 0.5n^{-1/5}$ and $h_\alpha = 0.75n^{-1/3}$ (for the plug-in method), $\epsilon = 0.001$ and $N = 10{,}000$

		Score-1		**Score-2**		**Plugin**		**MLE**	
	n	**mean**	***n* × var**	**mean**	***n* × var**	**mean**	***n* × var**	**mean**	***n* × var**
β	100	0.500212	0.364558	0.502247	0.410449	0.499562	0.245172	0.489690	0.307961
	500	0.499845	0.221484	0.499825	0.230178	0.498857	0.191857	0.499315	0.228335
	1000	0.499982	0.211608	0.500353	0.208102	0.499502	0.192223	0.499937	0.228420
	5000	0.499901	0.195294	0.499964	0.184807	0.500314	0.181421	0.499933	0.239898
	10,000	0.499988	0.191115	0.499985	0.172758	0.500120	0.172043	0.499994	0.227222
	20,000	0.500038	0.187616	0.500023	0.169762	0.500096	0.174197	0.499952	0.238400
α	100	0.511937	0.468415	0.509679	0.515638	0.495709	0.332949	0.523103	0.425614
	500	0.502258	0.293585	0.502506	0.287576	0.498932	0.254040	0.502514	0.304540
	1000	0.500839	0.284958	0.500616	0.262684	0.498385	0.270085	0.500937	0.300201
	5000	0.500345	0.262566	0.500316	0.244892	0.501597	0.241294	0.500270	0.303754
	10,000	0.500127	0.256983	0.500134	0.232973	0.501680	0.245993	0.500076	0.289905
	20,000	0.500020	0.250720	0.500042	0.230901	0.501660	0.244042	0.500101	0.302824

intercept are 0.257898 for the simple score method and 0.222984 for the efficient methods. Our results show slow convergence to these bounds.

Table 1 shows that the efficient score and the efficient plug-in methods perform reasonably well. A drawback of the plug-in method however is the long computing time for large sample sizes, whereas the computation for the MLE is fast even for the larger samples. Note moreover that the plug-in estimate is only asymptotically monotone whereas the MLE is monotone by definition. All our proposed estimates perform better than the classical MLE; the log likelihood for the MLE has moreover a rough behavior, with a larger chance that optimization algorithms might calculate a local maximizer instead of the global maximizer.

The performance of the score estimates is worse than the performance of the plug-in estimates for small sample sizes but increases considerably when the sample size increases. Although the asymptotic variance of the first score-estimator of Section 4.1 is larger than the (almost, determined by the truncation parameter ϵ) efficient variance, the results obtained with this method are noteworthy seen the fact that no smoothing is involved in this simple estimation technique.

Table 1 does not provide strong evidence of the $\sqrt{n}$-consistency of the classical MLE, but we conjecture that the MLE is indeed $\sqrt{n}$-consistent but not efficient. Considering the drawbacks of the classical MLE, we advise the use of the plug-in estimate for small sample sizes and the use of the score estimates, based on the MLE, for larger sample sizes, for estimating the parameter β_0. We finally suggest to estimate the parameter α_0 via the MLE corresponding to this β_0-estimate, avoiding in this way the bias problem for the kernel estimates of α_0.

6.1. *Bandwidth selection.* In this section, we discuss the bandwidth selection for the plug-in estimate. A similar idea can be used for the selection of the bandwidth used for the second estimate defined in Section 4.2. We define the optimal constant c_{opt} in $h = cn^{-1/5}$ as the minimizer of MSE,

$$c_{\mathrm{opt}} = \arg\min_c \mathrm{MSE}(c) = \arg\min_c E_{\beta_0}(\hat{\beta}_{n,h_c} - \beta_0)^2,$$

where $\hat{\beta}_{n,h_c}$ is the estimate obtained when the constant c is chosen in the estimation method. A picture of the Monte Carlo estimate of MSE as a function of c is shown for the plug-in method in Figure 7, where we estimated $\mathrm{MSE}(c)$ on a grid $c = 0.01, 0.05, 0.10, \ldots, 0.95$, for a sample size $n = 1000$ and truncation parameter $\epsilon = 0.001$ by a Monte Carlo experiment with $N = 1000$ simulation runs,

$$\widehat{\mathrm{MSE}(c)} = N^{-1} \sum_{j=1}^{N} (\hat{\beta}^j_{n,h_c} - \beta_0)^2, \tag{6.1}$$

where $\hat{\beta}^j_{n,h_c}$ is the estimate of β_0 in the jth simulation run, $j = 1, \ldots, N$.

Since F_0 and β_0 are unknown in practice, we cannot compute the actual MSE. We use the bootstrap method proposed by [26] to obtain an estimate of MSE. Our

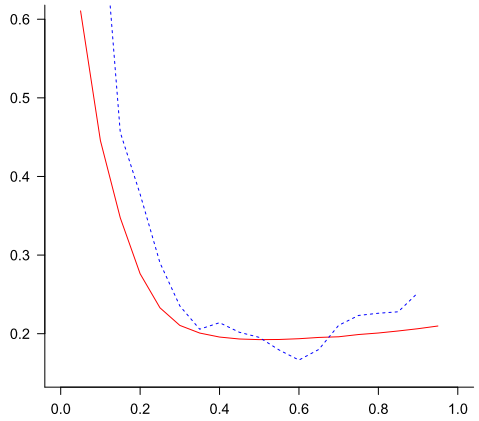

FIG. 7. *Estimated* MSE(c) *plot of* $\hat{\beta}_n$ *obtained from* $N = 1000$ *Monte Carlo simulations* (*red*, *solid*) *versus the bootstrap* MSE *for* $c_0 = 0.25$ (*blue*, *dashed*) *with* $B = 10{,}000$, $n = 1000$ *and* $\epsilon = 0.001$.

proposed estimate $F_{nh,\beta}$ of the distribution function F_0 satisfies the conditions of Theorem 3 in [26] and the consistency of the bootstrap is guaranteed. Note that it follows from [20] and [25] that naive bootstrapping, by resampling with replacement (T_i, X_i, Δ_i), or by generating bootstrap samples from the MLE, is inconsistent for reproducing the distribution of the MLE.

The method works as follows. We let $h_0 = c_0 n^{-1/5}$ be an initial choice of the bandwidth and calculate the plug-in estimates $\hat{\beta}_{n,h_0}$ and $F_{n,h_{c_0}}$ based on the original sample (T_i, X_i, Δ_i), $i = 1, \ldots, n$. We generate a bootstrap sample (X_i, T_i, Δ_i^*), $i = 1, \ldots, n$ where the (T_i, X_i) correspond to the (T_i, X_i) in the original sample and where the indicator Δ_i^* is generated from a Bernoulli distribution with probability $F_{n,h_0}(T_i - \hat{\beta}_{n,h_0} X_i)$, and next estimate $\hat{\beta}^*_{n,h_c}$ from this bootstrap sample. We repeat this B times and estimate MSE(c) by

$$\widehat{\mathrm{MSE}_B(c)} = B^{-1} \sum_{b=1}^{B} \bigl(\hat{\beta}^{*b}_{n,h_c} - \hat{\beta}_{n,h_{c_0}}\bigr)^2, \tag{6.2}$$

where $\hat{\beta}^{*b}_{n,h_c}$ is the bootstrap estimate in the bth bootstrap run. The optimal bandwidth $\hat{h}_{\mathrm{opt}} = \hat{c}_{\mathrm{opt}} n^{-1/5}$ where $\hat{c}_{\mathrm{opt}}$ is defined as the minimizer of $\widehat{\mathrm{MSE}_B(c)}$.

To analyze the behavior of the bootstrap method, we compared the Monte Carlo estimate of MSE, defined in (6.1), (based on $N = 1000$ samples of size $n = 1000$) to the bootstrap MSE defined in (6.2) (based on a single sample of size $n = 1000$) in Figure 7. The figure shows that the Monte Carlo MSE and the bootstrap MSE are in line, which illustrates the consistency of the method. The choice of the initial bandwidth does affect the size of the estimated MSE but not the behavior of the estimate, and we conclude that this bootstrap algorithm can be used to select an optimal bandwidth parameter in the described method above.

7. Discussion. In this paper, we propose a simple $\sqrt{n}$-consistent estimate for the finite dimensional regression parameter in the semiparametric current status linear regression model with unspecified error distribution. The estimate has an asymptotically normal limiting distribution but does not attain the efficiency bounds. We also consider two different methods to obtain $\sqrt{n}$-consistent and asymptotic normal estimates with an asymptotic variance that is arbitrarily close to the efficient variance. The first approach uses the MLE for the distribution function F for fixed β, the second approach does not depend on the behavior of this MLE but uses a kernel estimate for the distribution function. All proposed estimates are defined as a root of a score function as a function of β.

We introduced a truncation device to avoid theoretical and numerical difficulties caused by unbounded score functions. The truncation is carried out by considering a subsample of the data depending on the ϵ and $1-\epsilon$ quantiles of the distribution function estimate. Although our estimates do not attain the efficiency bound, the proposed method allows for easy computation of the estimates without the need for selecting a truncation sequence converging to zero. Achieving efficiency at the cost of additional computational complexities associated with smoothing procedures and truncation sequence selection results in only a small asymptotic efficiency gain and does not seem to improve the performance of our simple methods.

The estimates based on the efficient score function depending on the MLE for F_0 for fixed β have a slightly better performance than the estimates based on the smooth score function depending on the plug-in estimates for F_0 when the sample size is large. For small samples, none of the MLE-based estimates comes out as uniformly best.

APPENDIX

In this section, we include the derivation of the efficient information bound for the current status linear regression model. The proofs of the results given in Sections 3, 4 and 5 are deferred to the Supplementary Material [10].

A.1. Efficient information in the current status linear regression model. The density of one observation in the current status linear regression model is

$$p_{\beta,F}(t,x,\delta) = F\bigl(t-\beta' x\bigr)^{\delta}\bigl\{1-F\bigl(t-\beta' x\bigr)\bigr\}^{1-\delta} f_{T,X}(t,x).$$

We assume that the distribution of (T, X) does not depend on (β, F), which implies that the relevant part of the log likelihood is given by

$$l_n(\beta,F) = \sum_{i=1}^{n}\bigl[\Delta_i \log F\bigl(T_i-\beta' X_i\bigr) + (1-\Delta_i)\log\bigl\{1-F\bigl(T_i-\beta' X_i\bigr)\bigr\}\bigr].$$

If the distribution F is known (parametric case), the information for β is given by

$$I_P(\beta) = E\biggl(\biggl(\frac{\partial}{\partial\beta}\log p_{\beta,F}(T,X,\Delta)\biggr)'\biggl(\frac{\partial}{\partial\beta}\log p_{\beta,F}(T,X,\Delta)\biggr)\biggr).$$

Straightforward calculations yield

$$I_P(\beta)_{ij} = \int \frac{\mathbb{E}(X_i X_j | T - \beta' X = u)}{F(u)\{1 - F(u)\}} f(u)^2 f_{T-\beta' X}(u)\, du,$$

where $f = F'$ and where $f_{T-\beta' X}$ is the density of $T - \beta' X$. When F is unknown, we need to calculate the efficient score function. Let F and P_0 be the probability measures of ε and (T, X, Δ), respectively, and let $L_2^0(Q)$ be the Hilbert space of square integrable functions a with respect to the measure dQ satisfying $\int a\, dQ = 0$. The score operator $l_F : L_2^0(F) \mapsto L_2^0(P_0)$ is defined by

$$\begin{aligned}[l_F a](t, x, \delta) &= E\big(a(\varepsilon) | (T, X, \Delta) = (t, x, \delta)\big) \\ &= \frac{\delta \int_{-\infty}^{t-\beta' x} a(s)\, dF(s)}{F(t - \beta' x)} - \frac{(1 - \delta) \int_{-\infty}^{t-\beta' x} a(s)\, dF(s)}{1 - F(t - \beta' x)},\end{aligned}$$

with adjoint,

$$\big[l_F^* b\big](e) = E\big(b(T, X, \Delta) | \varepsilon = e\big).$$

The information for β in the semiparametric model is defined by

$$I(\beta) = E\big(\tilde{\ell}_{\beta, F}(T, X, \Delta)' \tilde{\ell}_{\beta, F}(T, X, \Delta)\big),$$

where $\tilde{\ell}_{\beta, F}(t, x, \delta)$ is the efficient score function defined by

$$\tilde{\ell}_{\beta, F}(t, x, \delta) = \ell_\beta(t, x, \delta) - [\ell_F a_*](t, x, \delta),$$

where

$$\ell_\beta(t, x, \delta) = \frac{\partial}{\partial \beta} \log p_{\beta, F}(t, x, \delta) = \frac{-\delta x f(t - \beta' x)}{F(t - \beta' x)} + \frac{(1 - \delta) x f(t - \beta' x)}{1 - F(t - \beta' x)},$$

and $\ell_F a_*$ satisfies

$$\ell_F^* \ell_F a_* = \ell_F^* \ell_\beta. \tag{A.1}$$

The efficient score $\tilde{\ell}_{\beta, F}$ can be interpreted as the residual of ℓ_β projected in the space spanned by $\ell_F a$ for $a \in L_2^0(F)$. Note that, as a consequence of (A.1), the efficient information is

$$I(\beta) = E\big(\tilde{\ell}_{\beta, F}(T, X, \Delta)' \ell_\beta(T, X, \Delta)\big).$$

To find a_*, we have to solve (A.1),

$$\begin{aligned}\ell_F^* \ell_F a_*(e) &= \int_e^\infty \frac{\phi(u)}{F(u)} f_{T-\beta' X}(u)\, du - \int_{-\infty}^e \frac{\phi(u)}{1 - F(u)} f_{T-\beta' X}(u)\, du \\ &= -\int_e^{+\infty} \frac{\mathbb{E}(X | T - \beta' X = u) f(u)}{1 - F(u)} f_{T-\beta' X}(u)\, du \\ &\quad + \int_{-\infty}^e \frac{\mathbb{E}(X | T - \beta' X = u) f(u)}{1 - F(u)} f_{T-\beta' X}(u)\, du \\ &= \ell_F^* \ell_\beta(e),\end{aligned} \tag{A.2}$$

where $\phi(t) = \int_{-\infty}^{t} a(s)\,dF(s)$. Equation (A.2) is satisfied with

$$\phi(u) = -\mathbb{E}(X|T - \beta'X = u)f(u).$$

Any a_* that satisfies the above equation satisfies (A.1) and we get

$$\tilde{\ell}_{\beta,F}(t,x,\delta) = \{E(X|T - \beta'X = t - \beta'x) - x\}f(t - \beta'x) \cdot \left\{\frac{\delta}{F(t - \beta'x)} - \frac{1-\delta}{1 - F(t - \beta'x)}\right\}$$

and

$$I(\beta)_{ij} = \int \frac{\mathrm{Cov}(X_i, X_j|T - \beta'X = u)}{F(u)\{1 - F(u)\}} f(u)^2 f_{T-\beta'X}(u)\,du. \tag{A.3}$$

Note that $I(\beta)^{-1} - I_P(\beta)^{-1}$ equals the minimal increase of the variance of an estimator for β based on an unknown F (semiparametric case) compared to the situation where F is known (parametric). In our simulation example, $I_P(\beta) = 26.3667$ and $I(\beta) = 6.5917$.

Acknowledgements. The authors want to thank Richard Nickl for useful comments and his reference to relevant theory in the book [9] and Fadoua Balabdaoui for incisive questions and references to relevant literature. We also thank the Associate Editor and two referees for their valuable remarks.

SUPPLEMENTARY MATERIAL

Supplement to "Current status linear regression" (DOI: 10.1214/17-AOS1589SUPP; .pdf). We give the proofs of the results stated in Sections 3, 4 and 5 of the manuscript.

REFERENCES

[1] CHEN, X., LINTON, O. and VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259

[2] COSSLETT, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51** 765–782. MR0712369

[3] COSSLETT, S. R. (1987). Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* **55** 559–585. MR0890854

[4] COSSLETT, S. R. (2007). Efficient estimation of semiparametric models by smoothed maximum likelihood. *Internat. Econom. Rev.* **48** 1245–1272. MR2375624

[5] DING, Y. and NAN, B. (2011). A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Statist.* **39** 3032–3061. MR3012400

[6] DOMINITZ, J. and SHERMAN, R. P. (2005). Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory* **21** 838–863. MR2189497

[7] FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42** 845–854. MR0872963

[8] FINKELSTEIN, D. M. and WOLFE, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945. MR0833140

[9] GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Univ. Press, New York. MR3588285

[10] GROENEBOOM, P. and HENDRICKX, K. (2018). Supplement to "Current status linear regression." DOI:10.1214/17-AOS1589SUPP.

[11] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators*, *Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293

[12] GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387. MR2589325

[13] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

[14] HAN, A. K. (1987). Nonparametric analysis of a generalized regression model. The maximum rank correlation estimator. *J. Econometrics* **35** 303–316. MR0903188

[15] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171

[16] HÖFFDING, W. (1940). Maszstabinvariante Korrelationstheorie. *Schr. Math. Inst. U. Inst. Angew. Math. Univ. Berlin* **5** 181–233. Translated in: *The Collected Works of Wassily Hoeffding* (N. I. Fisher and P. K. Sen, eds.). Springer, New York, 1994. MR0004426

[17] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975

[18] HUANG, J. and WELLNER, J. (1993). Regression models with interval censoring. In *Proceedings of the Kolmogorov Seminar*. Euler Mathematics Institute, St. Petersburg, Russia.

[19] KLEIN, R. W. and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61** 387–421. MR1209737

[20] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212

[21] LI, G. and ZHANG, C.-H. (1998). Linear regression with interval censored data. *Ann. Statist.* **26** 1306–1327. MR1647661

[22] MURPHY, S. A., VAN DER VAART, A. W. and WELLNER, J. A. (1999). Current status regression. *Math. Methods Statist.* **8** 407–425. MR1735473

[23] RABINOWITZ, D., TSIATIS, A. and ARAGON, J. (1995). Regression with interval-censored data. *Biometrika* **82** 501–513. MR1366277

[24] ROSSINI, A. J. and TSIATIS, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *J. Amer. Statist. Assoc.* **91** 713–721. MR1395738

[25] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880

[26] SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. MR3257405

[27] SHEN, X. (2000). Linear regression with current status data. *J. Amer. Statist. Assoc.* **95** 842–852. MR1804443

[28] SHERMAN, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61** 123–137. MR1201705

[29] SHIBOSKI, S. C. and JEWELL, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Amer. Statist. Assoc.* **87** 360–372.

[30] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

DELFT INSTITUTE OF APPLIED MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS
E-MAIL: P.Groeneboom@tudelft.nl

I-BIOSTAT
HASSELT UNIVERSITY
I-BIOSTAT, AGORALAAN
B3590 DIEPENBEEK
BELGIUM
E-MAIL: kim.hendrickx@uhasselt.be