

CURRENT STATUS LINEAR REGRESSION

BY PIET GROENEBOOM AND KIM HENDRICKX

Delft University of Technology and University Hasselt

We discuss estimators for the finite-dimensional regression parameter in the current status linear regression model. It is shown that, using a simple truncation device, one can construct \sqrt{n} -consistent and asymptotically normal estimates of the finite-dimensional regression parameter with an asymptotic covariance matrix that is arbitrarily close to the matrix of the information lower bound. We illustrate this with a simulation study and provide algorithms for computing the estimates and for selecting the bandwidth with a bootstrap method. The connection with results on the binary choice model in the econometric literature is also discussed.

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 sort of censoring. In this paper we focus on modeling a linear relationship when the response variable is subject to interval censoring type I, i.e. 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. [6] and [7]) and in HIV and AIDS studies (see e.g. [26]). Substantial literature has been devoted to regression models with current status data including the proportional hazard model studied in [13], the accelerated failure time model proposed by [21] and the proportional odds regression model of [22].

The regression model we want to study is the semi-parametric 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], [16] and [5]), where, however, the censoring variable T is degenerate, i.e. $P(T = 0) = 1$, and observations are of the type $(X_i, 1_{\{Y_i \leq 0\}})$. In the latter model, the scale is not identifiable, which one usually solves by adding a constraint on the parameter space such as setting the length of β or the the first coefficient equal to one.

The maximum likelihood estimator (MLE) of β_0 was proved to be consistent by [2] but nothing seems to be known about its asymptotic distribution, apart from its consistency and upper bounds for its rate of convergence. Since the log likelihood as a function of β , obtained by maximizing the log likelihood with respect to the distribution function F for fixed β and substituting this maximizer back into the likelihood, is not a smooth function of β , it is unclear if the MLE of β_0 is \sqrt{n} -consistent. [19] derived an $n^{-1/3}$ -rate for the MLE; we conjecture, based on simulation results that this can be strengthened to a $n^{-1/2}$ -rate, however the efficiency and limiting distribution of the MLE remains an open question.

Approaches to \sqrt{n} -consistent and efficient estimation of the regression parameters were considered by [16],[19], [18], [25] and [4] among others. For a derivation of the efficient information

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62-04

Keywords and phrases: MLE, SMLE, current status, linear regression, binary choice

$\tilde{\ell}_{\beta_0, F_0}^2$,

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

we refer to [3] for the binary choice model, and later to [14] and [19] for the current status regression model.

Often appearing in the literature about the current status regression model, is the condition that the support of the density of $T - \beta'_0 X$ is strictly contained in an interval D for all β and that F_0 stays strictly away from 0 and 1 on D (see e.g. [14],[19] and [25]). The drawback of this assumption is that we have no information about the whole distribution F_0 . This is also opposite to the usual conditions made for the current status model, where one 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 that arise if this condition is not made, such as numerical complications and unbounded score functions. Examples of truncation methods can be found in [4] and [16] among others.

In this paper we present some simple estimates for the finite dimensional regression parameter in the current status linear regression model. We construct a truncated log likelihood for the current status linear regression model and propose three estimates of the unknown error distribution in Section 2 that lead to three different estimates of the regression parameter. We introduce a simulation example to illustrate the methods and discuss the connection with previous results in the binary choice and current status regression models. In Section 3 we give the asymptotic behavior of the plug-in estimator which is the first estimator introduced in Section 2. We show that the 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. The estimation of an intercept term, that originates from the mean of the error distribution, is outlined in Section 4. Section 5 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. An Appendix is included in Section 6 with proofs of the results given in this paper.

2. The current status linear regression model. Let $(X_i, T_i, \Delta_i), i = 1, \dots, n$ be independent and identically distributed observations from $(X, T, \Delta) = (X, T, 1_{\{Y \leq T\}})$. We assume that Y is modeled as

$$(2.1) \quad Y = \beta'_0 X + \varepsilon,$$

where β_0 is a k -dimensional regression parameter and ε is an unobserved random error, independent of (X, T) with unknown distribution function F_0 . We assume that the distribution of (X, T) 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,

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

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,

$$(2.3) \quad l_n^{(\epsilon)}(\beta, F) = \int_{F(t - \beta'x) \in [\epsilon, 1 - \epsilon]} [\delta \log F(t - \beta'x) + (1 - \delta) \log\{1 - F(t - \beta'x)\}] d\mathbb{P}_n(t, x, \delta),$$

where $\epsilon \in (0, 1/2)$ is a truncation parameter. In principle one could choose $\epsilon = 0$, but this choice gives both theoretical and numerical difficulties and leads to an unbounded score function.

REMARK 2.1. If we use truncation, we have to prove that maximizing the log likelihood on a sub-interval still gives a consistent estimate of β_0 . This is done in Section 3. If one starts with the score equation or an estimate thereof, the solution sometimes suggested in the literature, is to add a constant c_n , tending to zero as $n \rightarrow \infty$, to the factor $F(t - \beta'x)\{1 - F(t - \beta'x)\}$ which inevitably will appear in the denominator. This is done in, e.g. [18]; similar ideas involving a sequence (c_n) are used in [16] and [4]. Picking a suitable sequence is more tricky, though, than just using the simple device in (2.3).

In what follows, we propose an estimation technique for β_0 based on three types of smoothed estimators for F_0 : (1) the plug-in estimator $F_{nh,\beta}$, (2) the smoothed maximum likelihood estimator (SMLE) $\tilde{F}_{nh,\beta}$ and (3) the penalized maximum likelihood estimator $\bar{F}_{n\lambda,\beta}$. For fixed β , we will construct an estimate \hat{F}_β of F_0 and then maximize the truncated log likelihood $l_n^{(\epsilon)}(\beta, \hat{F}_\beta)$ as a function of β to obtain an estimate of β_0 .

Throughout the paper, we illustrate our estimates by a simple simulated data example. Before we formulate our estimates, we first describe the simulation set-up. We consider the model,

$$(2.4) \quad Y_i = 0.5X_i + \varepsilon_i,$$

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 $E(\varepsilon) = 0.5$, our linear model contains an intercept,

$$E(Y_i|X_i = x_i) = 0.5 + 0.5x_i.$$

We next list our three methods for estimating β_0 .

2.1. *Method 1: The plug-in estimate $F_{nh,\beta}$.* Define

$$(2.5) \quad F_{nh,\beta}(t - \beta'x) = \frac{\int \delta K_h(t - \beta'x - u + \beta'y) d\mathbb{P}_n(u, y, \delta)}{\int K_h(t - \beta'x - u + \beta'y) d\mathbb{G}_n(u, y)},$$

where \mathbb{G}_n is the empirical distribution function of the pairs (T_i, X_i) , the probability measure of (T, X) will be denoted by G , and where K_h is a scaled version of a probability density function K given by,

$$K_h(\cdot) = h^{-1}K(h^{-1}(\cdot)) \quad \text{with bandwidth } h > 0,$$

satisfying condition (K.1). The triweight kernel is used in the simulation examples given in the remainder of the paper.

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

The plug-in estimates are not necessarily monotone but we prove in Theorem 3.2 that $F_{nh,\beta}$ is monotone with probability tending to one as $n \rightarrow \infty$ and $\beta \rightarrow \beta_0$. Another way of writing $F_{nh,\beta}$ is

in terms of ordinary sums. Let

$$(2.6) \quad g_{nh,1,\beta}(t - \beta'x) = \frac{1}{n} \sum_{j=1}^n \Delta_j K_h(t - \beta'x - T_j + \beta'X_j),$$

and,

$$(2.7) \quad g_{nh,\beta}(t - \beta'x) = \frac{1}{n} \sum_{j=1}^n K_h(t - \beta'x - T_j + \beta'X_j),$$

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. [12]). 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. We can define the corresponding score equation for β by,

$$\psi_n^{(\epsilon)}(\beta) = \mathbf{0},$$

where $\mathbf{0}$ is the k -dimensional vector with zeros as components and,

$$(2.8) \quad \psi_n^{(\epsilon)}(\beta) = \int_{F_{nh,\beta}(t - \beta'x) \in (\epsilon, 1 - \epsilon)} \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta'x) \frac{\delta - F_{nh,\beta}(t - \beta'x)}{F_{nh,\beta}(t - \beta'x) \{1 - F_{nh,\beta}(t - \beta'x)\}} d\mathbb{P}_n(t, x, \delta),$$

with

$$\frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta'x) = \left(\frac{\partial}{\partial \beta_1} F_{nh,\beta}(t - \beta'x), \dots, \frac{\partial}{\partial \beta_k} F_{nh,\beta}(t - \beta'x) \right),$$

The solution of the above equation is not necessarily unique, but by Rolle's theorem, the maximizer of (2.3) will approximately satisfy the score equation. Note, however, that there are some difficulties in defining the partial derivative of the truncated log likelihood with respect to β . For example, 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 ; so the score function is only an asymptotic representation of the partial derivatives of the truncated log likelihood.

The estimates of F_0 do not have a closed form expression in β ; we first have to estimate for fixed β the corresponding estimate of F_0 and next maximize the profile likelihood to obtain the estimate of β_0 . The estimates described in Methods 2 and 3 also follow the latter procedure.

2.2. Method 2: The smoothed maximum likelihood estimate (SMLE) $\tilde{F}_{nh,\beta}$. For fixed β , the MLE $\hat{F}_{n,\beta}$ of the distribution function of $Y - \beta'X$, is a piecewise constant function with jumps at a subset of $\{T_i - \beta'X_i : i = 1, \dots, n\}$ and can be characterized by the left derivative of the convex minorant of a cumulative sum diagram (see e.g. proposition 1.2 in [11] on p. 41). The SMLE is defined by

$$(2.9) \quad \tilde{F}_{nh,\beta}(t) = \int \mathbb{K} \left(\frac{t - x}{h} \right) d\hat{F}_{n,\beta}(x),$$

where,

$$\mathbb{K}(u) = \int_{-\infty}^u K(y) dy,$$

is the integrated kernel. Since the partial derivative w.r.t. β can, however, only be defined at values β where a slight change of the parameter β does not lead to a change of ordering of the values $T_i - \beta' X_i$, there are several (asymptotically equivalent) ways in which we can represent this partial derivative. We choose to represent it as the derivative of the toy estimator,

$$(2.10) \quad \tilde{F}_{nh,\beta}^{toy}(t - \beta'x) = \int \frac{\hat{F}_{n,\beta}(u - \beta'y)K_h(t - \beta'x - u + \beta'y)}{f_{T-\beta'X}(u - \beta'y)} dG(u, y)$$

So we use the representation

$$(2.11) \quad \begin{aligned} & \psi_n^{(\epsilon)}(\beta, \tilde{F}_{nh,\beta}) \\ &= \int_{\tilde{F}_{nh,\beta}(t-\beta'x) \in (\epsilon, 1-\epsilon)} \frac{\partial}{\partial \beta} \tilde{F}_{nh,\beta}(t - \beta'x) \frac{\delta - \tilde{F}_{nh,\beta}(t - \beta'x)}{\tilde{F}_{nh,\beta}(t - \beta'x)\{1 - \tilde{F}_{nh,\beta}(t - \beta'x)\}} d\mathbb{P}_n(t, x, \delta), \end{aligned}$$

where we estimate $\frac{\partial}{\partial \beta} \tilde{F}_{nh,\beta}(t - \beta'x)$ by

$$(2.12) \quad \begin{aligned} & \int \frac{(y-x)\hat{F}_{n,\beta}(u - \beta'y)K'_h(t - \beta'x - u + \beta'y)}{f_{T-\beta'X}(u - \beta'y)} d\mathbb{G}_n(u, y) \\ & + \int \frac{y\hat{F}_{n,\beta}(u - \beta'y)K_h(t - \beta'x - u + \beta'y)f'_{T-\beta'X}(u - \beta'y)}{f_{T-\beta'X}(u - \beta'y)^2} d\mathbb{G}_n(u, y), \end{aligned}$$

where x and y are k -dimensional vectors. Note however, that this expression is not used in the actual computation of $\tilde{F}_{nh,\hat{\beta}_n}$, since in this computation the log likelihood $l_n^{(\epsilon)}(\beta, \tilde{F}_{nh,\beta})$ is directly maximized w.r.t. the SMLE's $\tilde{F}_{nh,\beta}$, and no attempt is made to solve the score equation.

2.3. Method 3: The penalized estimate $\bar{F}_{n\lambda,\beta}$. Another interesting method to construct a smooth estimator of F_0 is via penalization (in analogy to the exposition in section 8.3 of [10]). For each β , define $\bar{F}_{n\lambda,\beta}$ as the *monotonic* minimizer F of

$$(2.13) \quad \int_{\bar{a}}^{\bar{b}} \{F(x) - \hat{F}_{n,\beta}(x)\}^2 dx + \lambda \int_{\bar{a}}^{\bar{b}} F'(x)^2 dx,$$

where $\hat{F}_{n,\beta}$ is again the MLE for fixed β and the interval $[\bar{a}, \bar{b}]$ is chosen in such a way that it contains all observations $T_i - \beta' X_i$. We let \bar{a} and \bar{b} depend on β and choose \bar{a} to be the smallest of the $T_i - \beta' X_i$ minus $\sqrt{\lambda}$, where λ is the penalty parameter in (2.13), and let \bar{b} be the largest order statistic of the $T_i - \beta' X_i$ plus $\sqrt{\lambda}$. Other choices are also possible. For fixed β the solution of the minimization problem, under the restriction that $F(\bar{a}) = 0$ and $F(\bar{b}) = 1$, is given by the solution of the second order differential equation (Euler's equation)

$$F''(x) = \{F(x) - \hat{F}_{n,\beta}(x)\}/\lambda,$$

with solution

$$(2.14) \quad \bar{F}_{n\lambda,\beta}(x) = \frac{1}{2\sqrt{\lambda}} \int_{\bar{a}}^{\bar{b}} e^{-|x-y|/\sqrt{\lambda}} \hat{F}_{n,\beta}(y) dy + c_1 e^{-(x-\bar{a})/\sqrt{\lambda}} + c_2 e^{-(\bar{b}-x)/\sqrt{\lambda}},$$

where c_1 and c_2 are determined by the boundary conditions $\bar{F}_{n\lambda,\beta}(\bar{a}) = 0$ and $\bar{F}_{n\lambda,\beta}(\bar{b}) = 1$. Hence we have

$$c_1 = -\frac{\int_{\bar{a}}^{\bar{b}} e^{-(y-\bar{a})/\sqrt{\lambda}} \hat{F}_{n,\beta}(y) dy + e^{-(\bar{b}-\bar{a})/\sqrt{\lambda}} \left\{ 2\sqrt{\lambda} - \int_{\bar{a}}^{\bar{b}} e^{-(\bar{b}-y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(y) dy \right\}}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}},$$

and

$$c_2 = \frac{e^{-(\bar{b}-\bar{a})/\sqrt{\lambda}} \int_{\bar{a}}^{\bar{b}} e^{-(y-\bar{a})/\sqrt{\lambda}} \hat{F}_{n,\beta}(y) dy + \left\{ 2\sqrt{\lambda} - \int_{\bar{a}}^{\bar{b}} e^{-(\bar{b}-y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(y) dy \right\}}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}}.$$

Using integration by parts, $\bar{F}_{n\lambda,\beta}(u)$ can easily be computed as a finite sum

$$\begin{aligned} \bar{F}_{n\lambda,\beta}(u) &= \hat{F}_{n,\beta}(u) - \frac{1}{2} e^{-(\bar{b}-u)/\sqrt{\lambda}} - \frac{1}{2} \int_{\bar{a}}^u e^{-(u-v)/\sqrt{\lambda}} d\hat{F}_{n,\beta}(v) + \frac{1}{2} \int_u^{\bar{b}} e^{-(v-u)/\sqrt{\lambda}} d\hat{F}_{n,\beta}(v) \\ &\quad + c_1 e^{-(u-\bar{a})/\sqrt{\lambda}} + c_2 e^{-(\bar{b}-u)/\sqrt{\lambda}}, \end{aligned}$$

where

$$\int_{\bar{a}}^u e^{-(u-v)/\sqrt{\lambda}} d\hat{F}_{n,\beta}(v) = \sum_{j:u_j < u} e^{-(u-u_j)/\sqrt{\lambda}} p_j$$

and

$$\int_u^{\bar{b}} e^{-(v-u)/\sqrt{\lambda}} d\hat{F}_{n,\beta}(v) = \sum_{j:u_j \geq u} e^{-(u_j-u)/\sqrt{\lambda}} p_j$$

are finite sums over the (weighted) masses p_j at the (usually few) points of jump u_j of the ordinary MLE of the distribution function of $Y - \beta'X$. The penalized estimate has again a convolution structure, like the SMLE, but this time the convolution is with a density of infinite support (the Laplace density), which, moreover, has a cusp at zero.

For the partial derivative of $\bar{F}_{n\lambda,\beta}$ w.r.t. β we use a similar representation as we used for the estimate, based on the SMLE. The penalized estimate is represented by

$$\begin{aligned} \bar{F}_{n\lambda,\beta}(t - \beta'x) &= \frac{1}{2\sqrt{\lambda}} \int_{\bar{a} < u - \beta'y < t - \beta'x} \frac{e^{-(t-\beta'x-u+\beta'y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(u - \beta'y)}{f_{T-\beta'X}(u - \beta'y)} dG(u, y) \\ &\quad + \frac{1}{2\sqrt{\lambda}} \int_{t - \beta'x < u - \beta'y < \bar{b}} \frac{e^{-(u-\beta'y-t+\beta'x)/\sqrt{\lambda}} \hat{F}_{n,\beta}(u - \beta'y)}{f_{T-\beta'X}(u - \beta'y)} dG(u, y) \\ &\quad + c_1 e^{-(t-\beta'x-\bar{a})/\sqrt{\lambda}} + c_2 e^{-(\bar{b}-t+\beta'x)/\sqrt{\lambda}}. \end{aligned}$$

We estimate $\frac{\partial}{\partial \beta} \bar{F}_{n\lambda,\beta}(t - \beta'x)$ by

$$\begin{aligned} & -\frac{1}{2\lambda} \int_{\bar{a} < u - \beta'y < t - \beta'x} \frac{(y-x) \hat{F}_{n,\beta}(u - \beta'y) e^{-(t-\beta'x-u+\beta'y)/\sqrt{\lambda}}}{f_{T-\beta'X}(u - \beta'y)} d\mathbb{G}_n(u, y) \\ & + \frac{1}{2\sqrt{\lambda}} \int_{t - \beta'x < u - \beta'y < \bar{b}} \frac{y \hat{F}_{n,\beta}(u - \beta'y) e^{-(u-\beta'y-t+\beta'x)/\sqrt{\lambda}} f'_{T-\beta'X}(u - \beta'y)}{f_{T-\beta'X}(u - \beta'y)^2} d\mathbb{G}_n(u, y) \\ (2.15) \quad & + \frac{x}{\sqrt{\lambda}} \left\{ c_1 e^{-(t-\beta'x-\bar{a})/\sqrt{\lambda}} - c_2 e^{-(\bar{b}-t+\beta'x)/\sqrt{\lambda}} \right\} + e^{-(t-\beta'x-\bar{a})/\sqrt{\lambda}} \frac{\partial}{\partial \beta} c_1 + e^{-(\bar{b}-t+\beta'x)/\sqrt{\lambda}} \frac{\partial}{\partial \beta} c_2, \end{aligned}$$

where x and y are again k -dimensional vectors, and where we estimate the derivatives of c_1 and c_2 in the same way by representing them as integrals w.r.t. dG_n . The expressions are given at the end of the Appendix.

A picture of the estimates described in Methods 1, 2 and 3 is shown in Figures 1 and 2 for our simulation example. The function $\beta \mapsto l_n^{(\epsilon)}(\beta, F_{nh,\beta})$ and its estimate of the partial derivatives $\psi_n^{(\epsilon)}(\beta)$ for the plug-in estimator are shown in Figure 3(a,b). The same is shown for the SMLE in Figure 3(c,d) and for the penalized estimate in Figure 4, where we replace $F_{nh,\beta}$ in (2.8) by $\tilde{F}_{nh,\beta}$ and $\bar{F}_{n\lambda,\beta}$ for the SMLE and the penalized estimate respectively and use the techniques of (2.12) and (2.15) to construct the estimate of the partial derivative of the truncated log likelihood w.r.t. β . It can be seen that the graphs are much smoother for the plug-in estimator. The curves are constructed by taking 100 equidistant evaluation points between 0.45 and 0.55 (the real value of β is 0.5) and just connecting the values of the estimator and its derivative respectively, at these points.

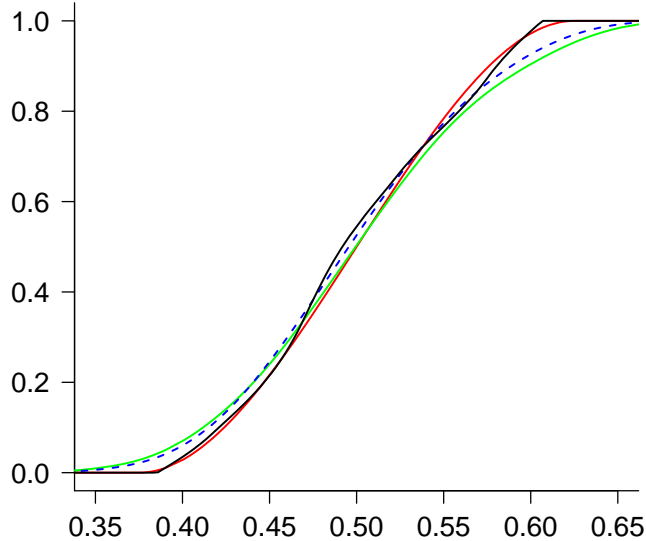


Fig 1: The real F_0 (red), the SMLE (dashed, blue), the plug-in estimate (green) and the penalized estimate (solid, black) for a sample of size $n = 1000$ with $\epsilon = 0.001$, $h = 0.5n^{-1/5}$ and $\sqrt{\lambda} = 0.125n^{-1/5}$.

In this paper, we give the asymptotic behavior of the estimator defined in Method 1. We prove that the plug-in estimator $\hat{\beta}_n$ of β_0 is consistent and asymptotically normal with an asymptotic covariance matrix that is arbitrarily close to the information lower bound. As a consequence of the truncation device used in our method, the information lower bound is not reached, but our simulation results suggest that the truncation effect is negligible in practice. The proofs are only given for the plug-in method but we conjecture that our proposed theoretical approach can be used in proving similar results for the SMLE and penalized estimates. We briefly sketch an outline of the proof for the other methods in Subsection 2.5.

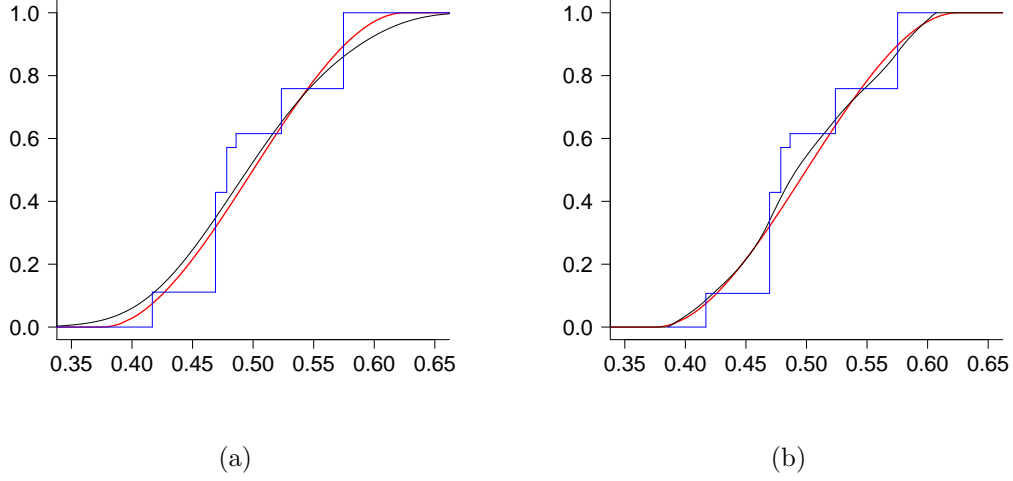


Fig 2: (a) The MLE (step function, blue) according to the SMLE (solid, black) and the real F_0 (red) and (b) The MLE (step function, blue) according to the penalized estimate (solid, black) and the real F_0 (red) .

2.4. *Result on the plug-in estimates and connection with other estimates.* Our main result is the following:

THEOREM 2.1. *Let $\beta_0 = (\beta_{0,1}, \dots, \beta_{0,k}) \in \mathbb{R}^k$ and let J_η denote the k -dimensional cube $\{\beta = (\beta_1, \dots, \beta_k) : \beta_i \in [\beta_{0,i} - \eta, \beta_{0,i} + \eta]\}$, for some $\eta > 0$. Let the distribution function F_0 be twice continuously differentiable on the interior of the support S of $f_0 = F_0'$, where S is an interval and let S be contained in the support of the density $f_{T-\beta'X}$, for each $\beta \in J_\eta$.*

Furthermore, let, for $\beta \in J_\eta$, the density $f_{T-\beta'X}(u)$ of $T - \beta'X$ and the conditional density $f_{X|T-\beta'X}(x|T - \beta'X = u)$ of X given $T - \beta'X$ be twice continuously differentiable functions w.r.t. u , except possibly at a finite number of points, and let, for $\beta \in J_\eta$, $f_{T-\beta'X}$ stay away from zero on the support of f_0 .

Finally, let $\beta \mapsto f_{T-\beta'X}(v)$ and $\beta \mapsto f_{X|T-\beta'X}(x|T - \beta'X = v)$ be continuous functions, for v and x in the definition domain of the functions and for $\beta \in J_\eta$, and let, for some $\epsilon \in (0, 1/2)$, $\hat{\beta}_n$ be the maximizer of

$$(2.16) \quad l_n^{(\epsilon)}(\beta) = \int_{F_{nh,\beta}(t-\beta'x) \in [\epsilon, 1-\epsilon]} [\delta \log F_{nh,\beta}(t - \beta'x) + (1 - \delta) \log\{1 - F_{nh,\beta}(t - \beta'x)\}] d\mathbb{P}_n(t, x, \delta),$$

where $F_{nh,\beta}$ is the plug-in estimate defined in (2.5), and let the function

$$\beta \mapsto \int_{F_0(t-\beta'x) \in [\epsilon, 1-\epsilon]} [F_0(t - \beta'_0 x) \log F_0(t - \beta'x) + (1 - F_0(t - \beta'_0 x)) \log\{1 - F_0(t - \beta'x)\}] dG(t, x), \quad \beta \in J_\eta,$$

have a unique maximum at $\beta = \beta_0$, where G is the distribution function of (T, X) .

Then, as $n \rightarrow \infty$, and $h \asymp n^{-1/5}$, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ converges in distribution to a k -dimensional normal distribution, with expectation zero and covariance matrix $I_\epsilon(\beta_0)^{-1}$, where $I_\epsilon(\beta_0)$ is the matrix

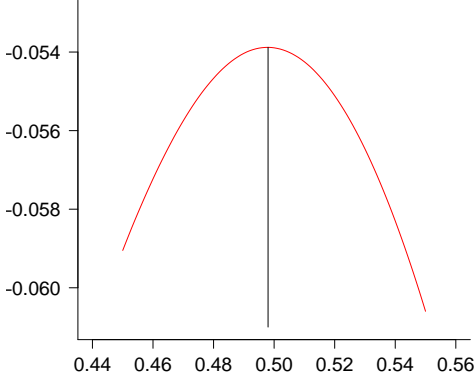
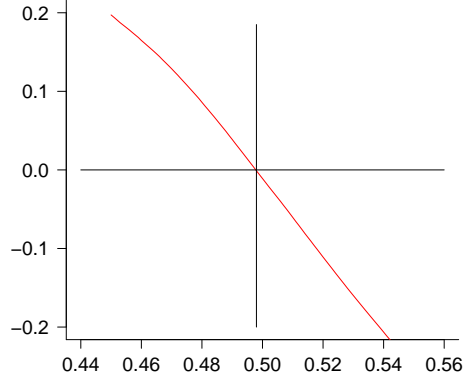
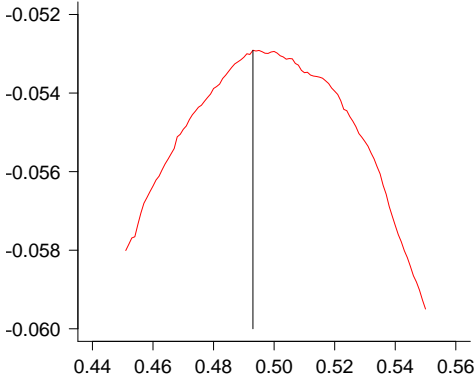
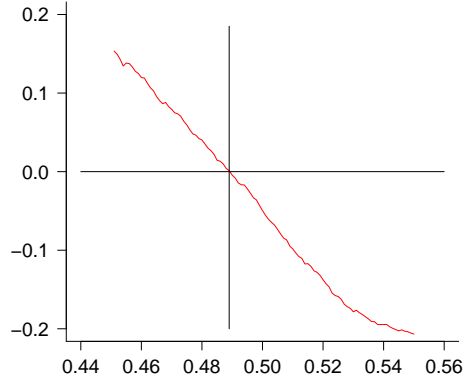
(a) $l_n^{(\epsilon)}(\beta, F_{nh,\beta})$ as function of β .(b) $\psi_n^{(\epsilon)}(\beta, F_{nh,\beta})$ as function of β .(c) $l_n^{(\epsilon)}(\beta, \tilde{F}_{nh,\beta})$ as function of β .(d) $\psi_n^{(\epsilon)}(\beta, \tilde{F}_{nh,\beta})$ as function of β .

Fig 3: The truncated log likelihood $l_n^{(\epsilon)}$ and its derivative w.r.t. β for the plug-in estimate $F_{nh,\beta}$ (a,b) and the SMLE $\tilde{F}_{nh,\beta}$ (c,d) for a sample of size $n = 1000$. The bandwidth $h = 0.5n^{-1/5}$ and $\epsilon = 0.001$. The vertical reference line in (a,c) indicates the location of the estimators $\hat{\beta}_{plugin,1} = 0.498$ and $\hat{\beta}_{SMLE,1} = 0.493$. The vertical reference line in (b,d) indicates the location of the zero of the score function $\hat{\beta}_{plugin,2} = 0.499$ and $\hat{\beta}_{SMLE,2} = 0.489$.

with elements

$$(2.17) \quad I_\epsilon(\beta_0)_{ij} = \int_{F_0(u) \in (\epsilon, 1-\epsilon)} \frac{\text{covar}(X_i, X_j | T - \beta_0'X = u)}{F_0(u)\{1 - F_0(u)\}} f_0(u)^2 f_{T-\beta_0'X}(u) du.$$

REMARK 2.2. [19] has (for the 1-dimensional case) the conditions that F_0 and $u \mapsto E_\beta(X|T - \beta X = u)$ are three times continuously differentiable instead of our condition of twice differentiability. Their asymptotic variance has the same form as ours, apart from the truncation parameter ϵ , but their variance has an implicit truncation since in [19] the integral only extends over a region where $F_0(1 - F_0)$ stays away from zero by assumption.

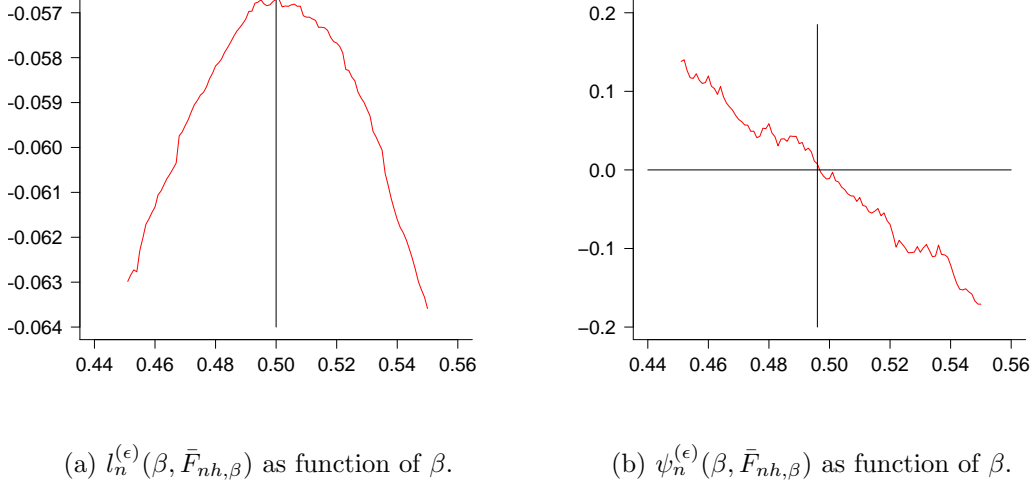


Fig 4: The truncated log likelihood $l_n^{(\epsilon)}$ and its derivative w.r.t. β for the penalized estimate $\bar{F}_{nh,\beta}$ for a sample of size $n = 1000$. The penalty parameter $\sqrt{\lambda} = 0.3n^{-1/5}$ and $\epsilon = 0.001$. The vertical reference line indicates the location of the estimators $\hat{\beta}_{penalized,1} = 0.500$ in (a) and the location of the zero of the score function $\hat{\beta}_{penalized,2} = 0.496$ in (b).

REMARK 2.3. The authors of [18] propose an estimation equation for β , derived from an inequality on the conditional covariance between X and Δ conditional on $T - \beta'X$, which has a U-statistics representation. Results on U-statistics are used to prove asymptotic normality and efficiency of their estimator. Note that the plug-in method also suggests the use of U-statistics; by straightforward calculations, we can write the score function defined in (2.8) as

$$\begin{aligned}
\psi_n^{(\epsilon)}(\beta) &= \frac{1}{n} \sum_{i=1}^n \frac{\{\Delta_i - F_{nh,\beta}(T_i - \beta'X_i)\} \frac{\partial}{\partial \beta} F_{nh,\beta}(T_i - \beta'X_i)}{F_{nh,\beta}(T_i - \beta'X_i) \{1 - F_{nh,\beta}(T_i - \beta'X_i)\}}, \\
&= \left\{ \frac{1}{n} \sum_{i=1}^n \Delta_i \frac{\frac{1}{n} \sum_{j \neq i} \Delta_j (X_j - X_i) K'_h(T_i - \beta'X_i - T_j + \beta'X_j)}{g_{nh,1,\beta}(T_i - \beta'X_i)} \right. \\
&\quad + \frac{1}{n} \sum_{i=1}^n (1 - \Delta_i) \frac{\frac{1}{n} \sum_{j \neq i} (1 - \Delta_j) (X_j - X_i) K'_h(T_i - \beta'X_i - T_j + \beta'X_j)}{g_{nh,0,\beta}(T_i - \beta'X_i)} \\
&\quad \left. - \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{n} \sum_{j \neq i} (X_j - X_i) K'_h(T_i - \beta'X_i - T_j + \beta'X_j)}{g_{nh,\beta}(T_i - \beta'X_i)} \right\} 1_{(\epsilon, 1-\epsilon)} \{F_{nh,\beta}(T_i - \beta'X_i)\},
\end{aligned}
\tag{2.18}$$

where $g_{nh,0,\beta} = g_{nh,\beta} - g_{nh,1,\beta}$, see (2.6) and (2.7). Each of the three terms on the right-hand side of (2.18) can be rewritten in terms of a scaled second order U-statistics. Obviously, a proof based on U-statistics would not be generalizable to proofs for the SMLE and the penalized estimators and requires in addition, lengthy and tedious calculations which are avoided in the current approach for proving Theorem 2.1. For this reasons, we do not further examine the results on U-statistics.

REMARK 2.4. The choice of the bandwidth h (in Methods 1 and 2) and the penalty parameter λ (in Method 3) is crucial to obtain good estimates of β_0 . We propose the bandwidth $h \asymp n^{-1/5}$ which is the usual bandwidth for ordinary second order kernels; a natural choice for λ is to choose $\lambda \asymp$

$n^{-2/5}$. Unfortunately, various advices are given in the literature on what smoothing parameters one should use. [16] has fourth order kernels and uses bandwidths between the orders $n^{-1/6}$ and $n^{-1/8}$. 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 $h \asymp n^{-1/5}$. [19] considers, for the current status model with a 1-dimensional regression parameter β , a penalized maximum likelihood estimator defined as the maximizer of

$$\sum_{i=1}^n \{\Delta_i \log F(T_i - \beta X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta X_i)\} - \lambda_n \int F''(u)^2 du,$$

where

$$1/\lambda_n = O_p\left(n^{2/5}\right), \quad \lambda_n^2 = o_p\left(n^{-1/2}\right).$$

Translated into bandwidth choice (using $h_n \asymp \sqrt{\lambda_n}$), the conditions correspond to: $n^{-1/5} \leq h \ll n^{-1/8}$, suggesting that their conditions do allow the choice $h \asymp n^{-1/5}$ or $\lambda \asymp n^{-2/5}$. [18] states that the bandwidth choice $h \asymp n^{-1/3}$ in their estimation procedure will yield an estimate of β_0 that is not efficient, which strengthens our conjecture that the MLE is \sqrt{n} -consistent but inefficient. It is however unclear to us, how to choose the bandwidth in [18].

2.5. Road map of the proof of Theorem 2.1. The older proofs of a result of this type always used second derivative calculations. As convincingly argued in [28], proofs of this type should only use first derivatives and that is indeed what we do; our proof follows more or less the structure of the proof of Theorem 25.54 in [28]. This means that we first prove a Donsker property for the functions representing the score function, see Lemma 3.2 and next prove (in Lemma 3.3), for the case that β is 1-dimensional, that the integral w.r.t. $d\mathbb{P}_n$ of this score function is

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

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 (2.17). Combining these two results gives Theorem 2.1.

Very crucial in this proof is Lemma 3.1, which gives L_2 -bounds on the distance of the estimate $F_{nh,\beta}$ of the distribution function of $Y - \beta X$ to its limit for fixed β (part (i)) and on the L_2 -distance between the first derivative $\frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x)$ of the estimate w.r.t. the parameter β and its limit (part(ii)). 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.

Another crucial tool is a result of [1], telling us that the functions considered belong to the right entropy class for applying the equicontinuity lemma of empirical process theory, using that, if $h \asymp n^{-1/5}$, the second derivative of the estimator is not consistent, but is in fact square integrable. Here we use a result in [8] on L_2 -bounds for derivatives of density estimates.

A similar road map can be followed for the other estimates we introduced above. Lemma 3.1 will again play a pivotal role. The partial derivative of the estimators, as defined here, are less smooth than the derivative of the plug-in estimator, but the global behavior is exactly the same, as can be seen in Figures 3 and 4. Also, the smoothness of these curves increases considerably with increasing sample size.

As an example (again for simplicity for the case that β is 1-dimensional), the SMLE will first be represented by the toy estimator, defined in (2.10), and the partial derivative w.r.t. β is represented by

$$\begin{aligned} \frac{\partial}{\partial \beta} \tilde{F}_{nh,\beta}(t - \beta x) &= \frac{\partial}{\partial \beta} \int \frac{\hat{F}_{n,\beta}(u - \beta y) K_h(t - \beta x - u + \beta y)}{f_{T-\beta X}(u - \beta y)} dG(u, y) \\ &= \int \frac{(y - x) \hat{F}_{n,\beta}(u - \beta y) K'_h(t - \beta x - u + \beta y)}{f_{T-\beta X}(u - \beta y)} dG(u, y) \\ &\quad + \int \frac{y \hat{F}_{n,\beta}(u - \beta y) K_h(t - \beta x - u + \beta y) f'_{T-\beta X}(u - \beta y)}{f_{T-\beta X}(u - \beta y)^2} dG(u, y), \end{aligned}$$

not using the (finite) number of points where the derivative cannot be defined. The only difference with (2.12) is that we integrate w.r.t. dG instead of dG_n . The partial derivative is next analyzed applying smooth functional theory in a somewhat similar way as the toy estimator is used in the proof of the asymptotic normality and efficiency of the estimate $\sqrt{n} \int \kappa_{F_0} d(\hat{F}_n - F_0)$ on the right-hand side of (10.20) in Section 10.2 of [10]. The penalized estimate can be treated along similar lines. We chose to prove the result in detail only for the plug-in estimate, since proving the details for the other estimates would take this paper out of bounds.

3. Asymptotic behavior of the plug-in estimator. In this section, we give the consistency, asymptotic monotonicity and the asymptotic distribution of the plug-in estimator $\hat{\beta}_n$ defined in Theorem 2.1.

THEOREM 3.1. *Let the conditions of Theorem 2.1 be satisfied. Then $\hat{\beta}_n$ is a consistent estimator of β_0 .*

The proof of Theorem 3.1 is given in the Appendix and was inspired by the arguments in section 4 of part II of [11], which were motivated by [15]. To prove the asymptotic monotonicity of the plug-in estimator, we follow the arguments of Theorem 3.3 of [9]. We get the following result.

THEOREM 3.2. *Let the conditions of Theorem 2.1 be satisfied, then we have on each interval I contained in the support of f_0 :*

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

For simplicity, we derive Theorem 2.1 for the one-dimensional case, using the results below. Let $\hat{\beta}_n$ be the maximizer of the truncated log likelihood, defined in (2.16), but now with $k = 1$. The partial derivative of $F_{nh,\beta}(t - \beta x)$, given by (2.5), w.r.t. β has the following form:

$$(3.1) \quad \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) = \frac{\int (y - x) \{ \delta - F_{nh,\beta}(t - \beta x) \} K'_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(t - \beta x)},$$

where $g_{nh,\beta}(t - \beta x)$ is defined in (2.7). Moreover, for the partial derivative of the truncated log likelihood $l_n^{(\epsilon)}$ with respect to β , defined by (2.8), we get

$$(3.2) \quad \psi_n^{(\epsilon)}(\hat{\beta}_n) = 0.$$

Although $\hat{\beta}_n$ might not be a unique solution of (3.2), every solution of the score equation will satisfy the results stated in the remainder of this section. The proof of Theorem 2.1 follows by combining Lemma 3.2 and Lemma 3.3 given below, proofs of both can be found in the Appendix. A crucial role in these proofs is played by the following two lemmas, also proved in the Appendix.

LEMMA 3.1. *Let the conditions of Theorem 2.1 be satisfied and let $k = 1$.*

(i) *Let the function F_β be defined by*

$$(3.3) \quad F_\beta(t - \beta x) = \int F_0(t - \beta x + (\beta - \beta_0)y) f_{X|T-\beta X}(y|t - \beta x) dy.$$

Then,

$$(3.4) \quad \int_{\epsilon < F_{nh,\beta}(t-\beta x) < 1-\epsilon} \{F_{nh,\beta}(t - \beta x) - F_\beta(t - \beta x)\}^2 dG(t, x) = O_p\left(\frac{1}{nh}\right) + O_p(h^4),$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$.

(ii) *Let the function a_β be defined by,*

$$(3.5) \quad \begin{aligned} a_\beta(t - \beta x) = & \int (y - x) f_0(t - \beta x + (\beta - \beta_0)y) f_{X|T-\beta X}(y|T - \beta X = t - \beta x) dy \\ & + f_{T-\beta X}(t - \beta x)^{-1} \int (y - x) \{F_0(t - \beta_0 x + (\beta - \beta_0)(y - x)) - F_\beta(t - \beta x)\} \\ & \cdot \frac{\partial}{\partial v} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|T - \beta X = v)\} \Big|_{v=t-\beta x} dy. \end{aligned}$$

Then,

$$(3.6) \quad \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) - a_\beta(t - \beta x) \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh^3}\right) + O_p(h^2).$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$.

(iii) *The results of (i) and (ii) remain valid when dG in (3.4) or (3.6) is replaced by dG_n .*

Under our conditions on the bandwidth h , which we assume to be of the usual order $n^{-1/5}$, Lemma 3.1 tells us that (1) the L_2 -distance between the estimate $F_{nh,\beta}$ and F_β , using dG or dG_n as dominating measure, and $t - \beta x$ as argument, is of order $n^{-2/5}$ and (2) the L_2 -distance between the derivative of $F_{nh,\beta}$ and the function a_β , using dG or dG_n as measure, is of order $n^{-1/5}$. In both cases we restrict the integration interval for the L_2 -distance to the interval $\epsilon < F_{nh,\beta} < 1 - \epsilon$.

This allows us, for example, to state that, by an application of the Cauchy-Schwarz inequality,

$$(3.7) \quad \begin{aligned} & \sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ a_\beta(t - \beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) \right\} \\ & \cdot \left\{ \frac{F_0(t - \beta_0 x) - F_{nh,\beta}(t - \beta x)}{F_{nh,\beta}(t - \beta x) \{1 - F_{nh,\beta}(t - \beta x)\}} \right\} dG_n(t, x) \\ & = O_p\left(n^{-1/10}\right) + o_p\left(\sqrt{n}(\beta - \beta_0)\right), \quad \beta \rightarrow \beta_0, \end{aligned}$$

which is an essential step in the proof of the ‘‘Donsker property’’ given in the next Lemma.

REMARK 3.1. Under the assumption that the functions $f_{T-\beta'X}(u)$ and $f_{X|T-\beta'X}(x|T-\beta'X = u)$ are three times continuously differentiable functions w.r.t. u for $\beta \in J_\eta$, the results of Lemma 3.1(i) can be improved to

$$\int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) - a_\beta(t - \beta x) \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh^3}\right) + O_p(h^4).$$

All bandwidth choices in the range $n^{-1/5} \lesssim h \ll n^{-1/8}$ guarantee that the integral in (3.7) is $o_p(1 + \sqrt{n}(\beta - \beta_0))$.

LEMMA 3.2 (Donsker property). *Let the conditions of Theorem 2.1 be satisfied and let $[a, b]$ be the support of f_0 . Let $\eta > 0$ be chosen in such a way that $a_1(\beta) = F_\beta^{-1}(\epsilon) > a$, $b_1(\beta) = F_\beta^{-1}(1-\epsilon) < b$ and F_β is bounded away from 0 and 1 on $[a_1(\beta), b_1(\beta)]$, for each $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ where F_β is defined in (3.3).*

Then, for $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$,

$$(3.8) \quad \sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} a_\beta(t-\beta x) \frac{\delta - F_{nh,\beta}(t-\beta x)}{F_{nh,\beta}(t-\beta x)\{1 - F_{nh,\beta}(t-\beta x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta)$$

is asymptotically normal, with expectation zero and asymptotic variance

$$(3.9) \quad \int_{F_\beta(t-\beta x) \in (\epsilon, 1-\epsilon)} a_\beta(t-\beta x)^2 \frac{F_0(t-\beta_0 x) - 2F_0(t-\beta_0 x)F_\beta(t-\beta x) + F_\beta(t-\beta x)^2}{F_\beta(t-\beta x)^2\{1 - F_\beta(t-\beta x)\}^2} dG(t, x) - \left\{ \int_{F_\beta(t-\beta x) \in (\epsilon, 1-\epsilon)} a_\beta(t-\beta x) \frac{F_0(t-\beta_0 x) - F_\beta(t-\beta x)}{F_\beta(t-\beta x)\{1 - F_\beta(t-\beta x)\}} dG(t, x) \right\}^2,$$

where a_β is defined in (3.5).

REMARK 3.2. Note that the variance (3.9) given in Lemma 3.2 is for $\beta = \beta_0$ given by:

$$(3.10) \quad I_\epsilon(\beta_0) \stackrel{\text{def}}{=} \int_{F_0(t-\beta_0 x) \in (\epsilon, 1-\epsilon)} \frac{a_{\beta_0}(t-\beta_0 x)^2}{F_0(t-\beta_0 x)\{1 - F_0(t-\beta_0 x)\}} dG(t, x),$$

where

$$a_{\beta_0}(t-\beta_0 x) = f_0(t-\beta_0 x)E_{\beta_0}\{X-x|T-\beta_0 X = t-\beta_0 x\}.$$

When $\epsilon = 0$, $I_\epsilon(\beta_0)$ equals the information bound for the current status linear regression model (see e.g [14]).

Next we use that $\hat{\beta}_n$ is a maximum likelihood estimator and in particular a solution of the score equation (3.2).

LEMMA 3.3. *Let the conditions of Theorem 2.1 be satisfied. Then,*

$$(3.11) \quad \sqrt{n} \int_{F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x) \in (\epsilon, 1-\epsilon)} a_{\hat{\beta}_n}(t-\hat{\beta}_n x) \frac{\delta - F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)}{F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)\{1 - F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)\}} dP_0(t, x, \delta)$$

is asymptotically equivalent to $-\sqrt{n}I_\epsilon(\beta_0)(\hat{\beta}_n - \beta_0)$, where $I_\epsilon(\beta_0)$ is given by (3.10), and

$$(3.12) \quad \sqrt{n} \int_{F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x) \in (\epsilon, 1-\epsilon)} a_{\hat{\beta}_n}(t-\hat{\beta}_n x) \frac{\delta - F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)}{F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)\{1 - F_{nh,\hat{\beta}_n}(t-\hat{\beta}_n x)\}} d\mathbb{P}_n(t, x, \delta) = o_p\left(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)\right).$$

Theorem 2.1 then follows by combining Lemma 3.2 and Lemma 3.3. The higher-dimensional extension is straightforward. Perhaps the easiest method is to use the Cramér-Wold device and consider linear combinations of the components of $\hat{\beta}_n$ on which we apply the preceding arguments.

In Section 4 we also need the following representation of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

THEOREM 3.3. *Let the conditions of Theorem 2.1 be satisfied. Then,*

$$\begin{aligned} & \sqrt{n}(\hat{\beta}_n - \beta_0) \\ &= n^{-1/2} I_\epsilon(\beta_0)^{-1} \sum_{i=1}^n f_0(T_i - \beta_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta_0 X_i) - X_i\} \\ & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta_0 X_i)}{F_0(T_i - \beta_0 X_i) \{1 - F_0(T_i - \beta_0 X_i)\}} 1_{(\epsilon, 1-\epsilon)} \{F_0(T_i - \beta_0 X_i)\} + o_p(1). \end{aligned}$$

REMARK 3.3. Lemma 3.3 and Theorem 3.3 show two sides of the coin, so to speak, of the proof of the asymptotic normality and efficiency of $\hat{\beta}_n$. Using the property that $\hat{\beta}_n$ is a maximum likelihood estimator, we get that:

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1-\epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{\delta - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= o_p\left(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)\right). \end{aligned}$$

and therefore that the leading asymptotic behavior of the integral w.r.t. $d(\mathbb{P}_n - P_0)$ is given by

$$\sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1-\epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{\delta - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} dP_0(t, x, \delta),$$

which is asymptotically equivalent to $-\sqrt{n} I_\epsilon(\beta_0)(\hat{\beta}_n - \beta_0)$, using the Donsker property.

Theorem 3.3 shows that the leading term of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ is given by a sum of independent random variables, involving the efficient score function

$$f_0(t - \beta_0 x) \{E_{\beta_0}(X | T - \beta_0 X = t - \beta_0 x) - x\} \frac{\delta - F_0(t - \beta_0 x)}{F_0(t - \beta_0 x) \{1 - F_0(t - \beta_0 x)\}} 1_{(\epsilon, 1-\epsilon)} \{F_0(t - \beta_0 x)\}.$$

In this case the integral w.r.t. dP_0 is zero, whereas in the preceding representation the integral w.r.t. $d\mathbb{P}_n$ was (sufficiently close to) zero.

The last representation plays a crucial role in determining the variance of smooth functionals, of which the intercept is an example. The representation of Theorem 3.3 also indicates that the U-statistics representation, which can be used for proofs of the asymptotic behavior of the plug-in estimator, does not give the most natural approach to the proof of asymptotic normality and efficiency of $\hat{\beta}_n$.

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

$$(4.1) \quad \alpha = \int u dF_0(u).$$

We can take the plug-in estimate $\hat{\beta}_n$ of β_0 , by using a bandwidth of order $n^{-1/5}$ and the maximum likelihood procedure, as before. However, in estimating α , as defined by (4.1), 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], p. 1253.

We have the following result of which the proof can be found in the Appendix.

THEOREM 4.1. *Let the conditions of Theorem 2.1 be satisfied, and let $\hat{\beta}_n$ be the k -dimensional estimate of β_0 as obtained by the maximum likelihood procedure, described in Theorem 2.1, 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

$$(4.2) \quad \sigma^2 \stackrel{\text{def}}{=} a(\beta_0)' I_\epsilon(\beta_0)^{-1} a(\beta_0) + \int \frac{F_0(v)\{1 - F_0(v)\}}{f_{T-\beta_0 X}(v)} dv,$$

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

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

and $I_\epsilon(\beta_0)$ is as in Theorem 2.1.

REMARK 4.1. We chose 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 4.2. 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 (4.2) is familiar from current status theory, see e.g. (10.7), p. 287 of [10].

REMARK 4.3. Instead of considering the plug-in estimate, we could also consider the SMLE, described in Subsection 2.2. After having determined an estimate $\hat{\beta}$ in this way, we next estimate α by

$$(4.3) \quad \hat{\alpha} = \int x d\hat{F}_{n, \hat{\beta}}(x),$$

where $\hat{F}_{n, \hat{\beta}}$ is the MLE and not the SMLE corresponding to this $\hat{\beta}$ to avoid bias, in accordance with re-estimating $F_{nh, \hat{\beta}_n}$ under a different bandwidth as done in Theorem 4.1. The MLE is roughly comparable with a kernel estimate with a locally adaptive bandwidth of order $n^{-1/3}$, and the bias is vanishing in the local asymptotic distribution.

5. Computation and simulations. The computation of our estimates is relatively straightforward in all cases. For the plug-in estimate, we simply compute the estimate as a ratio of two kernel estimators for fixed β . Next we maximize the log likelihood over β , using Brent's optimization procedure for non-linear optimization in one dimension. For dimensions larger than one, Broyden's method can be used. For the SMLE and the penalized estimator, we first compute the MLE for fixed β by the so-called "pool adjacent violators" algorithm for computing the convex minorant of the so-called "cusum diagram", consisting of the points (0, 0) and

$$\left(i, \sum_{j=1}^i \Delta_j \right), \quad i = 1, \dots, n,$$

for the observations (T_i, X_i, Δ_i) , where the ordering is according to the ordering of the $T_i - \beta'X_i$, for fixed β : $T_1 - \beta'X_1 \leq T_2 - \beta'X_2, \dots$. After the MLE is computed for fixed β , we can compute either the SMLE or the penalized MLE. Since the SMLE is in fact a weighted sum of the few masses of the MLE, computation is very fast. The same is true for the penalized MLE, where only the constants c_1 and c_2 have to be determined from two linear equations, once we have determined the MLE. The estimate of β_0 is then again determined by an optimization algorithm.

Some results from the simulations of our model are given in Tables 1 and 2. Table 1 (resp. Table 2) contains the mean value of the estimate, averaged over $N = 1000$ 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 for different sample sizes n and different truncation parameters ϵ . We chose the bandwidth $h = 0.5n^{-1/5}$ for the plug-in and SMLE methods and the penalty parameter $\sqrt{\lambda} = 0.125n^{-1/5}$ for the penalized method based on an investigation of the mean squared error (MSE) for different choices of c in $h = cn^{-1/5}$ and $\sqrt{\lambda} = cn^{-1/5}$. Details on how to choose the bandwidth in practice are given in Subsection 5.1. From Tables 1 and 2, it is seen that the estimates obtained without truncation, i.e. $\epsilon = 0$, are not favored above those obtained with our proposed truncation device. This can be explained by the instable behavior of the likelihood at the boundary. 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, 0.153859 for $\epsilon = 0.0001$, 0.158699 for $\epsilon = 0.001$ and 0.17596 for $\epsilon = 0.1$. Our results show slow convergence to these bounds. We advise to use a truncation parameter ϵ of 0.001 or smaller in practice. Tables 1 and 2 show that all proposed 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 SMLE and the penalized methods is fast even for the larger samples. Also added to Tables 1 and 2 is the performance of the MLE, we see that the variance stabilizes when the sample sizes increases but remains larger than the corresponding values of the variances for the other methods, strengthening our belief that the MLE is \sqrt{n} -consistent but not efficient.

TABLE 1

The mean value of the estimate and n times the variance of the estimate of β_0 for different methods, $h = 0.5n^{-1/5}$, $\sqrt{\lambda} = 0.125n^{-1/5}$ and $N = 1000$.

		Plug-in		SMLE		Penalized		MLE.	
ϵ	n	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$
0.001	100	0.499562	0.245172	0.498230	0.206338	0.496977	0.211718	0.473455	0.186408
	500	0.498857	0.191857	0.498862	0.199093	0.498812	0.201979	0.498111	0.306398
	1000	0.499502	0.192223	0.499433	0.197447	0.499432	0.197220	0.499528	0.296097
	5000	0.500314	0.181421	0.500249	0.185238	0.500151	0.187605	0.500264	0.219442
	10000	0.500120	0.172043	0.500119	0.174606	0.500118	0.179449	0.500159	0.233758
	20000	0.500096	0.174197	0.500090	0.170635	0.500034	0.183646	0.500050	0.236070
0	100	0.499587	0.244887	0.498174	0.205513	0.496973	0.211676	0.473455	0.186408
	500	0.498857	0.191591	0.498855	0.198557	0.498825	0.201961	0.498111	0.306398
	1000	0.499498	0.191797	0.499414	0.197101	0.499438	0.196697	0.499528	0.296097
	5000	0.500310	0.180752	0.500235	0.184497	0.500163	0.187588	0.500264	0.219442
	10000	0.5001187	0.171181	0.500110	0.174371	0.500115	0.180662	0.500159	0.233758
	20000	0.500094	0.173041	0.500091	0.169200	0.500035	0.182125	0.500050	0.236070

5.1. *Bandwidth selection.* We define the optimal constant c_{opt} in $h = cn^{-1/5}$ and $\sqrt{\lambda} = cn^{-1/5}$ as the minimizer of MSE ,

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

TABLE 2

The mean value of the estimate and n times the variance of the estimate of α_0 for different methods, $h = 0.5n^{-1/5}$ and $\sqrt{\lambda} = 0.125n^{-1/5}$ are used in the estimation of β_0 , $h_\alpha = 0.75n^{-1/3}$ is used for the estimation of α_0 , $N = 1000$.

ϵ	n	Plug-in		SMLE		Penalized		MLE.	
		$mean(\hat{\alpha}_n)$	$nvar(\hat{\alpha}_n)$	$mean(\hat{\alpha}_n)$	$nvar(\hat{\alpha}_n)$	$mean(\hat{\alpha}_n)$	$nvar(\hat{\alpha}_n)$	$mean(\hat{\alpha}_n)$	$nvar(\hat{\alpha}_n)$
0.001	100	0.495729	0.332986	0.510547	0.303242	0.511939	0.307928	0.534931	0.302948
	500	0.498932	0.254040	0.503088	0.265176	0.503083	0.265234	0.503871	0.374480
	1000	0.498385	0.270085	0.501460	0.274942	0.501440	0.279836	0.501292	0.381873
	5000	0.501594	0.241959	0.499929	0.248425	0.500031	0.249143	0.499905	0.280939
	10000	0.501679	0.246909	0.499926	0.251921	0.499930	0.251029	0.499882	0.310525
	20000	0.501658	0.245014	0.499930	0.240735	0.499985	0.257535	0.499969	0.313966
0	100	0.495709	0.332949	0.510465	0.301574	0.511950	0.307784	0.534931	0.302948
	500	0.498930	0.257908	0.503064	0.263495	0.503062	0.265623	0.503871	0.374480
	1000	0.498389	0.269702	0.501494	0.275547	0.501428	0.279513	0.501292	0.381873
	5000	0.501597	0.241294	0.499941	0.257552	0.500021	0.249149	0.499905	0.280939
	10000	0.501680	0.245993	0.499936	0.250405	0.499932	0.251583	0.499882	0.310525
	20000	0.501660	0.244042	0.499929	0.238863	0.499983	0.255136	0.499969	0.313966

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 three methods in Figure 5, where we estimated $MSE(c)$ on a grid $c = 0.01, 0.05, 0.10, \dots, 0.95$, for a sample size $n = 1000$ and truncation parameter $\epsilon = 0.001$ by a Monte Carlo experiment with $N = 1000$ simulation runs,

$$(5.1) \quad \widehat{MSE}(c) = N^{-1} \sum_{j=1}^N (\hat{\beta}_{n,h_c}^j - \beta_0)^2,$$

where $\hat{\beta}_{n,h_c}^j$ is the estimate of β_0 in the j -th simulation run, $j = 1, \dots, N$. It is seen from this picture that the optimal bandwidth for the penalized method is clearly smaller than for the other two methods, and the plug-in estimate seems to have the largest optimal bandwidth.

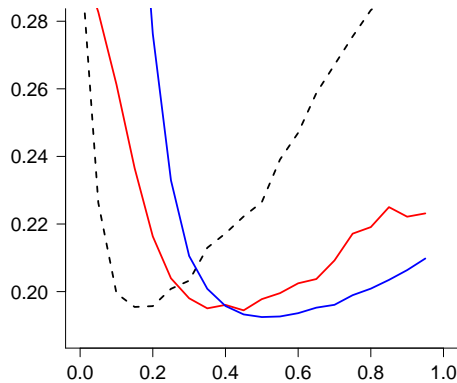


Fig 5: Estimated MSE of $\hat{\beta}_n$ for the plug-in estimate (blue), the SMLE (red) and the penalized estimate (dashed), with $n = 1000$ and $\epsilon = 0.001$, and using $N = 1000$ Monte Carlo samples.

Since F_0 and β_0 are unknown in practice, we cannot compute MSE . We use the bootstrap method

proposed by [24] to obtain an estimate of MSE . Our proposed estimates of the distribution function F_0 satisfy the conditions of Theorem 3 in [24] and the consistency of the bootstrap is guaranteed. Note that it follows from [17] and [23] 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 (for the SMLE, the others methods work similar). We let $h_0 = c_0 n^{-1/5}$ be an initial choice of the bandwidth and calculate the SMLE estimates $\hat{\beta}_{n,h_0}$ and $\tilde{F}_{n,h_{c_0}}$ based on the original sample $(X_i, T_i, \Delta_i), i = 1, \dots, n$. We generate a bootstrap sample $(X_i, T_i, \Delta_i^*), i = 1, \dots, n$ where the (X_i, T_i) correspond to the (X_i, T_i) in the original sample and where the indicator Δ_i^* is generated from a Bernoulli distribution with probability $\tilde{F}_{n,h_0}(T_i - \hat{\beta}_{n,h_0} X_i)$ and next estimate $\hat{\beta}_{n,h_c}^*$ from this bootstrap sample. We repeat this B times and estimate $MSE(c)$ by,

$$(5.2) \quad \widehat{MSE}_B(c) = B^{-1} \sum_{b=1}^B (\hat{\beta}_{n,h_c}^{*b} - \hat{\beta}_{n,h_{c_0}})^2,$$

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

To analyze the behavior of the bootstrap method, we compared the Monte Carlo estimate of MSE , defined in (5.1), (based on $N = 1000$ samples of size $n = 1000$) to the bootstrap MSE defined in (5.2) (based on a single sample of size $n = 1000$) in Figure 6 for the plug-in and SMLE method and in Figure 7 for the penalized method. The figures show 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 effect 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 or penalty parameter in the described methods above.

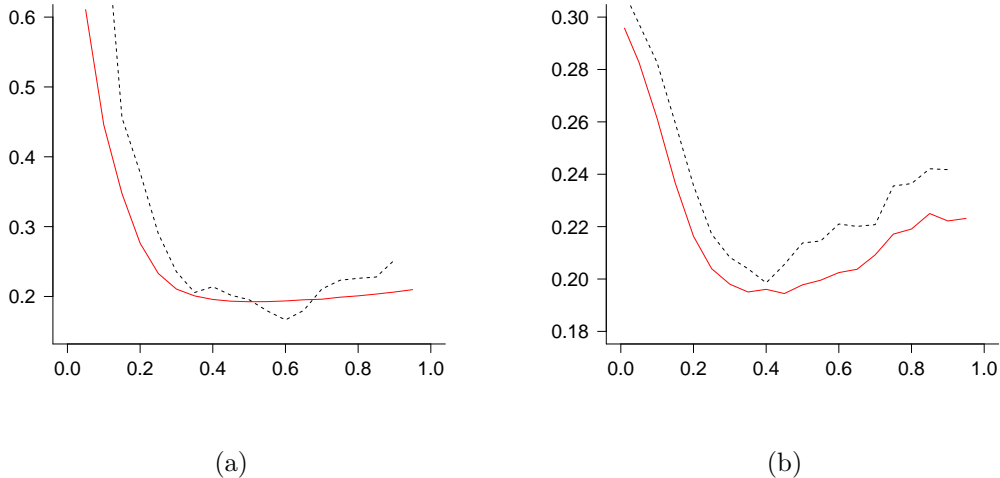


Fig 6: Estimated $MSE(c)$ plot of $\hat{\beta}_n$ obtained from 1000 Monte Carlo simulations (red, solid) versus the bootstrap MSE for $c_0 = 0.25$ (dashed, black) with $B = 10000$, $n = 1000$ and $\epsilon = 0.001$ for (a) the plug-in method and (b) the SMLE method.

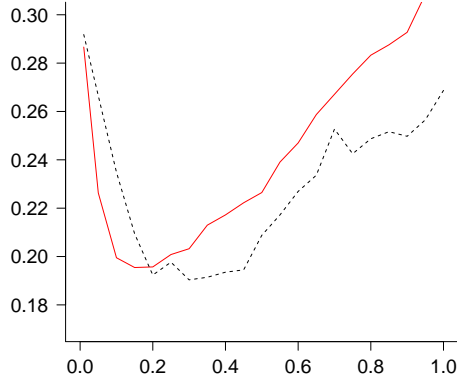


Fig 7: Estimated $MSE(c)$ plot of $\hat{\beta}_n$ for the penalized method, obtained from 1000 Monte Carlo simulations (red, solid) versus the bootstrap MSE for $c_0 = 0.125$ (dashed, black) with $B = 10000$, $n = 1000$ and $\epsilon = 0.001$.

6. Appendix. In this section we give the proofs of the results of the previous sections.

PROOF OF THEOREM 3.1. Consider maximizing

$$(6.1) \quad l_n^{(\epsilon)}(F_{nh,\beta}) = \int_{\epsilon < F_{nh,\beta}(t - \beta'x) < 1 - \epsilon} \{\delta \log F_{nh,\beta}(t - \beta'x) + (1 - \delta) \log \{1 - F_{nh,\beta}(t - \beta'x)\}\} d\mathbb{P}_n(t, x, \delta),$$

over β , where $F_{nh,\beta}$ is defined by (2.5). Let $\hat{\beta}_n$ be the value maximizing (6.1) via $F_{nh,\hat{\beta}_n}$. Since $\hat{\beta}_n$ is the maximizer of $l_n^{(\epsilon)}(F_{nh,\beta})$, we must have, for each $\lambda \in (0, 1)$:

$$l_n^{(\epsilon)}\left((1 - \lambda)F_{nh,\hat{\beta}_n} + \lambda F_{nh,\beta_0}\right) - l_n^{(\epsilon)}\left(F_{nh,\hat{\beta}_n}\right) \leq 0.$$

Hence,

$$\limsup_{\lambda \downarrow 0} \lambda^{-1} \left\{ l_n^{(\epsilon)}\left((1 - \lambda)F_{nh,\hat{\beta}_n} + \lambda F_{nh,\beta_0}\right) - l_n^{(\epsilon)}\left(F_{nh,\hat{\beta}_n}\right) \right\} \leq 0.$$

Note that

$$(6.2) \quad \begin{aligned} & \limsup_{\lambda \downarrow 0} \lambda^{-1} \left\{ l_n^{(\epsilon)}\left((1 - \lambda)F_{nh,\hat{\beta}_n} + \lambda F_{nh,\beta_0}\right) - l_n^{(\epsilon)}\left(F_{nh,\hat{\beta}_n}\right) \right\} \\ &= \limsup_{\lambda \downarrow 0} \lambda^{-1} \left\{ \int_{\epsilon < (1 - \lambda)F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh,\beta_0}(t - \beta'_0 x) < 1 - \epsilon} \right. \\ & \quad \left. \left\{ \delta \log \left\{ (1 - \lambda)F_{nh,\hat{\beta}_n}(t - x\hat{\beta}'_n x) + \lambda F_{nh,\beta_0}(t - \beta'_0 x) \right\} \right. \right. \\ & \quad \left. \left. + (1 - \delta) \log \left\{ 1 - (1 - \lambda)F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) - \lambda F_{nh,\beta_0}(t - \beta'_0 x) \right\} \right\} d\mathbb{P}_n(t, x, \delta) \right. \\ & \quad \left. - \int_{\epsilon < F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) < 1 - \epsilon} \left\{ \delta \log F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x) + (1 - \delta) \log \{1 - F_{nh,\hat{\beta}_n}(t - \hat{\beta}'_n x)\} \right\} d\mathbb{P}_n(t, x, \delta) \right\}. \end{aligned}$$

Since the areas of integration do not coincide for both integrals, we cannot combine the difference between these two integrals into one integral. In what follows, we first show that the limits for $\lambda \downarrow 0$, considered over the regions where both terms disagree are zero with probability one. Let A_n be defined by

$$A_n = \{(t, x) : F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \leq \epsilon, (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh, \beta_0}(t - \beta'_0 x) > \epsilon\}.$$

and let $A_{n, \lambda}$ be defined by

$$A_{n, \lambda} = \{(t, x) : F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \leq \epsilon - \lambda, (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh, \beta_0}(t - \beta'_0 x) > \epsilon\}.$$

Then, for $(t, x) \in A_{n, \lambda}$,

$$\begin{aligned} & \lambda \left\{ F_{nh, \beta_0}(t - \beta'_0 x) - \mathbb{F}_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \right\} \\ &= (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh, \beta_0}(t - \beta'_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) > \lambda, \end{aligned}$$

and hence

$$F_{nh, \beta_0}(t - \beta'_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) > 1,$$

which cannot occur. So we find that the set A_n is equal to the set $A'_{n, \lambda}$, defined by

$$A'_{n, \lambda} = \{(t, x) : \epsilon - \lambda < F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \leq \epsilon, (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh, \beta_0}(t - \beta'_0 x) > \epsilon\}.$$

Now note that, because of the preceding relation,

$$\begin{aligned} & \frac{1}{\lambda} \int_{\epsilon - \lambda < F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) \leq \epsilon} \left\{ \delta \log \left\{ (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) + \lambda F_{nh, \beta_0}(t - \beta'_0 x) \right\} \right. \\ & \quad \left. + (1 - \delta) \log \left\{ 1 - (1 - \lambda)F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - \lambda F_{nh, \beta_0}(t - \beta'_0 x) \right\} \right\} d\mathbb{P}_n \\ &= (n\lambda)^{-1} \sum_{i: F_{nh, \hat{\beta}_n}(T_i - \hat{\beta}'_n X_i) \in (\epsilon - \lambda, \epsilon]} \left\{ \Delta_i \log \left\{ (1 - \lambda)F_{nh, \hat{\beta}_n}(T_i - \hat{\beta}'_n X_i) + \lambda F_{nh, \beta_0}(T_i - \beta'_0 X_i) \right\} \right. \\ & \quad \left. + (1 - \Delta_i) \log \left\{ 1 - (1 - \lambda)F_{nh, \hat{\beta}_n}(T_i - \hat{\beta}'_n X_i) - \lambda F_{nh, \beta_0}(T_i - \beta'_0 X_i) \right\} \right\} \\ & \longrightarrow 0, \quad \lambda \downarrow 0. \end{aligned}$$

In fact, there is, with probability one, a (random) $\bar{\lambda} > 0$ such that

$$\{i \in \{1, \dots, n\} : F_{nh, \hat{\beta}_n}(T_i - \hat{\beta}'_n X_i) \in (\epsilon - \lambda, \epsilon]\} = \emptyset \text{ for } \lambda < \bar{\lambda}$$

and therefore the sum in the expression above will be zero for $\lambda < \bar{\lambda}$. We can deal in a similar way with the other situations in which parts of the integration regions in the two integrals in (6.2) do not match. Hence,

$$\begin{aligned} & \limsup_{\lambda \downarrow 0} \lambda^{-1} \left\{ l^{(\epsilon)}((1 - \lambda)F_{nh, \hat{\beta}_n} + \lambda F_{nh, \beta_0}) - l^{(\epsilon)}(F_{nh, \hat{\beta}_n}) \right\} \\ &= \int_{\epsilon < F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) < 1 - \epsilon} \left\{ \delta \frac{F_{nh, \beta_0}(t - \beta'_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)} \right. \\ & \quad \left. + (1 - \delta) \frac{1 - F_{nh, \beta_0}(t - \beta'_0 x) - (1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x))}{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)} \right\} d\mathbb{P}_n(t, x, \delta), \end{aligned}$$

and we get,

$$\begin{aligned} & \int_{\epsilon < F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) < 1 - \epsilon} \left\{ \delta \frac{F_{nh, \beta_0}(t - \beta'_0 x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)} + (1 - \delta) \frac{1 - F_{nh, \beta_0}(t - \beta'_0 x)}{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)} \right\} d\mathbb{P}_n(t, x, \delta) \\ & \leq \int_{\epsilon < F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) < 1 - \epsilon} d\mathbb{G}_n(t, x), \end{aligned}$$

We assume that $\hat{\beta}_n$ is contained in the cube J_η , and hence the sequence $(\hat{\beta}_n)$ has a subsequence $(\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega))$, converging to an element β . So, if $\hat{\beta}_{n_k} = \hat{\beta}_{n_k}(\omega) \rightarrow \beta$, we get,

$$F_{n_k h, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) \rightarrow F_\beta(t - \beta' x) \stackrel{\text{def}}{=} \int F_0(t - \beta' x + (\beta - \beta_0)' y) f_{X|T-\beta' X}(y|t - \beta' x) dy,$$

and since $F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)$ and $F_\beta(t - \beta' x)$ stay away from 0 and 1, we get from the convergence of $F_{n_k h, \beta_0}(t - \beta'_0 x; \omega)$ to $F_0(t - \beta'_0 x)$ and $F_{n_k h, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x; \omega)$ to $F_\beta(t - \beta' x)$ that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \int_{\epsilon < F_{n_k h, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x) < 1 - \epsilon} \left\{ \delta \frac{F_{n_k h, \beta_0}(t - \beta'_0 x; \omega)}{F_{n_k h, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x; \omega)} + (1 - \delta) \frac{F_{n_k h, \beta_0}(t - \beta'_0 x; \omega)}{1 - F_{n_k h, \hat{\beta}_{n_k}}(t - \hat{\beta}'_{n_k} x; \omega)} \right\} d\mathbb{P}_{n_k}(t, x, \delta) \\ & = \int_{\epsilon < F_\beta(t - \beta' x) < 1 - \epsilon} \left\{ \frac{F_0(t - \beta'_0 x)^2}{F_\beta(t - \beta' x)} + \frac{(1 - F_0(t - \beta'_0 x))^2}{1 - F_\beta(t - \beta' x)} \right\} dG(t, x) \\ & \leq \int_{\epsilon < F_\beta(t - \beta' x) < 1 - \epsilon} dG(t, x). \end{aligned}$$

This can only happen if $F_\beta(t - \beta' x) = F_0(t - \beta'_0 x)$ for all (t, x) such that $\epsilon < F_\beta(t - \beta' x) < 1 - \epsilon$ (see p. 78 of [11]). Since the argument can be repeated for each subsequence, this gives the consistency of $\hat{\beta}_n$. \square

To prove Theorem 3.2, we start with some results on the consistency of the kernel estimators $g_{nh, \hat{\beta}_n}, g_{nh1, \hat{\beta}_n}$ and their derivatives $g'_{nh, \hat{\beta}_n}$ and $g'_{nh1, \hat{\beta}_n}$ stated in the next Lemma.

LEMMA 6.1. *Let the conditions of Theorem 2.1 be satisfied and let $k = 1$, then we have,*

$$\begin{aligned} g_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) - g_{nh, \beta_0}(t - \beta_0 x) \\ = (\hat{\beta}_n - \beta_0) \frac{\partial}{\partial t} \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} + o_p(\hat{\beta}_n - \beta_0), \end{aligned}$$

$$\begin{aligned} g_{nh1, \hat{\beta}_n}(t - \hat{\beta}_n x) - g_{nh1, \beta_0}(t - \beta_0 x) \\ = (\hat{\beta}_n - \beta_0) \left\{ F_0(t - \beta_0 x) \frac{\partial}{\partial t} \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} \right. \\ \left. + f_0(t - \beta_0 x) \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} \right\} + o_p(\hat{\beta}_n - \beta_0), \end{aligned}$$

$$\begin{aligned} g'_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) - g'_{nh, \beta_0}(t - \beta_0 x) \\ = (\hat{\beta}_n - \beta_0) \frac{\partial^2}{\partial t^2} \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} + O_p(\hat{\beta}_n - \beta_0), \end{aligned}$$

and,

$$\begin{aligned} g'_{nh1, \hat{\beta}_n}(t - \hat{\beta}_n x) - g'_{nh1, \beta_0}(t - \beta_0 x) \\ = (\hat{\beta}_n - \beta_0) \left\{ f'_0(t - \beta_0 x) E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x) \right. \\ \left. + 2f_0(t - \beta_0 x) \frac{\partial}{\partial t} \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} \right. \\ \left. + F_0(t - \beta_0 x) \frac{\partial^2}{\partial t^2} \{E\{X|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} \right\} + O_p(\hat{\beta}_n - \beta_0). \end{aligned}$$

PROOF OF LEMMA 6.1. For the first expression, we obtain,

$$\begin{aligned} g_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) - g_{nh, \beta_0}(t - \beta_0 x) \\ = \int \left\{ K_h(t - \hat{\beta}_n x - u - \hat{\beta}_n y) - K_h(t - \beta_0 x - u - \beta_0 y) \right\} d\mathbb{G}_n(u, y) \\ = (\hat{\beta}_n - \beta_0) \int (y - x) K'_h(t - \beta_0 x - u + \beta_0 y) d\mathbb{G}_n(u, y) + o_p(\hat{\beta}_n - \beta_0) \\ = (\hat{\beta}_n - \beta_0) \int (y - x) K'_h(t - \beta_0 x - u + \beta_0 y) dG(u, y) \\ + (\hat{\beta}_n - \beta_0) \int (y - x) K'_h(t - \beta_0 x - u + \beta_0 y) d(\mathbb{G}_n - G)(u, y) + o_p(\hat{\beta}_n - \beta_0) \end{aligned}$$

Using integration by parts and standard kernel calculations, we get

$$\int (y - x) K'_h(t - \beta_0 x - u + \beta_0 y) dG(u, y) = \frac{\partial}{\partial t} \{E(X|T - \beta_0 X = t - \beta_0 x)g(t - \beta_0 x)\} + O(h).$$

Also,

$$\int (y - x) K'_h(t - \beta_0 x - u + \beta_0 y) d(\mathbb{G}_n - G)(u, y) = O_p\left(\frac{1}{nh^3}\right).$$

This proves the first result. The second result of Lemma 6.1 follows analogously.

For the third expression, we have using successive integration by parts,

$$\begin{aligned}
& g'_{n,\hat{\beta}_n}(t - \hat{\beta}_n x) - g'_{n,\beta_0}(t - \beta_0 x) \\
&= \int \left\{ K'_h(t - \hat{\beta}_n x - u - \hat{\beta}_n y) - K'_h(t - \beta_0 x - u - \beta_0 y) \right\} d\mathbb{G}_n(u, y) \\
&= (\hat{\beta}_n - \beta_0) \int (y - x) K''_h(t - \beta_0 x - u + \beta_0 y) dG(u, y) \\
&\quad + (\hat{\beta}_n - \beta_0) \int (y - x) K''_h(t - \beta_0 x - u + \beta_0 y) d(G_n - G)(u, y) + o_p(\hat{\beta}_n - \beta_0)
\end{aligned}$$

Applying integration by parts twice and using standard kernel calculations, we get

$$\int (y - x) K''_h(t - \beta_0 x - u + \beta_0 y) dG(u, y) = \frac{\partial^2}{\partial t^2} \{E(X|T - \beta_0 X = t - \beta_0 x)g(t - \beta_0 x)\} + O(1).$$

Also,

$$\int (y - x) K''_h(t - \beta_0 x - u + \beta_0 y) d(\mathbb{G}_n - G)(u, y) = O_p(1),$$

since for $h \asymp n^{-1/5}$, the order of the variance of this term is given by the order of,

$$\int (y - x)^2 (K''_h)^2(t - \beta_0 x - u + \beta_0 y) dG(u, y) = O\left(\frac{1}{nh^5}\right),$$

The third result of the lemma follows, to obtain the last result we apply similar calculations. \square

LEMMA 6.2. *Under the assumptions of Theorem 2.1,*

$$\sup_{t,x} |g_{n,\hat{\beta}_n}(t - \hat{\beta}_n x) - g(t - \beta_0 x)| = o_p(1), \quad \sup_{t,x} |g'_{n,\hat{\beta}_n}(t - \hat{\beta}_n x) - g'(t - \beta_0 x)| = o_p(1),$$

$$\sup_{t,x} |g_{n1,\hat{\beta}_n}(t - \hat{\beta}_n x) - g_1(t - \beta_0 x)| = o_p(1), \quad \text{and} \quad \sup_{t,x} |g'_{n1,\hat{\beta}_n}(t - \hat{\beta}_n x) - g'_1(t - \beta_0 x)| = o_p(1).$$

PROOF. For the first result, we have,

$$\begin{aligned}
\sup_{t,x} |g_{n,\hat{\beta}_n}(t - \hat{\beta}_n x) - g(t - \beta_0 x)| &\leq \sup_{t,x} |g_{n,\hat{\beta}_n}(t - \hat{\beta}_n x) - g_{n,\beta_0}(t - \beta_0 x)| \\
&\quad + \sup_{t,x} |g_{n,\beta_0}(t - \beta_0 x) - g(t - \beta_0 x)| \\
&\lesssim |\hat{\beta}_n - \beta_0| \sup_{t,x} \left| \frac{\partial}{\partial t} \{E\{X - x|T - \beta_0 X = t - \beta_0 x\}g(t - \beta_0 x)\} \right| \\
&\quad + \sup_{t,x} |g_{n,\beta_0}(t - \beta_0 x) - g(t - \beta_0 x)|.
\end{aligned}$$

The result now follows from the convergence of $\hat{\beta}_n$ to β_0 , the (two times) continuous differentiability of the functions $f_{T-\beta'X}$ and $f_{X|T-\beta'X}$ and Lemma A.2 in [9]. The other results are proved similarly. \square

PROOF OF THEOREM 3.2. The proof of Theorem 3.2 follows by similar arguments as the proof of Theorem 3.3 in [9] on p. 26, using the result of Lemma 6.2. \square

PROOF OF LEMMA 3.1. Part(i)

Recall that,

$$F_{nh,\beta}(t - \beta x) = \frac{g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)}$$

where

$$g_{nh,1,\beta}(t - \beta x) = \int \delta K_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta),$$

and

$$g_{nh,\beta}(t - \beta x) = \int K_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta).$$

Moreover,

$$F_\beta(t - \beta x) = \int F_0(t - \beta_0 x + (\beta - \beta_0)(y - x)) f_{X|T-\beta X}(y|t - \beta x) dy.$$

We first investigate the bias part.

$$\begin{aligned} \mathbb{E}g_{nh,1,\beta}(t - \beta x) &= \int F_0(u - \beta_0 y) K_h(t - \beta x - u + \beta y) dG(u, y) \\ &= \int F_0(v + (\beta - \beta_0)y) K_h(t - \beta x - v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dy dv \\ &= \int F_0(t - \beta x + (\beta - \beta_0)y - hw) K(w) f_{T-\beta X}(t - \beta x - hw) f_{X|T-\beta X}(y|t - \beta x - hw) dy dw \\ &= f_{T-\beta X}(t - \beta x) \int F_0(t - \beta_0 x + (\beta - \beta_0)(y - x)) f_{X|T-\beta X}(y|t - \beta x) dy + O(h^2), \end{aligned}$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, due to the assumptions of Theorem 2.1. In a similar way, we get

$$\mathbb{E}g_{nh,\beta}(t - \beta x) = f_{T-\beta X}(t - \beta x) + O(h^2),$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval. So we find:

$$\frac{\mathbb{E}g_{nh,1,\beta}(t - \beta x)}{\mathbb{E}g_{nh,\beta}(t - \beta x)} = F_\beta(t - \beta x) + O(h^2).$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, such that $\mathbb{E}g_{nh,1,\beta}(t - \beta x)$ stays away from zero.

So we obtain

$$\begin{aligned} &F_{nh,\beta}(t - \beta x) - F_\beta(t - \beta x) \\ &= \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} + \mathbb{E}g_{nh,1,\beta}(t - \beta x) \frac{\mathbb{E}g_{nh,\beta}(t - \beta x) - g_{nh,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x) \mathbb{E}g_{nh,\beta}(t - \beta x)} + O(h^2), \end{aligned}$$

and

$$\begin{aligned}
& \{F_{nh,\beta}(t - \beta x) - F_\beta(t - \beta x)\}^2 \\
& \leq 3 \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 + 3 \left\{ \mathbb{E}g_{nh,1,\beta}(t - \beta x) \frac{\mathbb{E}g_{nh,\beta}(t - \beta x) - g_{nh,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)\mathbb{E}g_{nh,\beta}(t - \beta x)} \right\}^2 \\
(6.3) \quad & + O(h^4).
\end{aligned}$$

uniformly in $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and t, x varying over a finite interval, such that $\mathbb{E}g_{nh,1,\beta}(t - \beta x)$ stays away from zero.

Since $\eta > 0$ is chosen in such a way that $a_1(\beta) = F_\beta^{-1}(\epsilon) > a$, $b_1(\beta) = F_\beta^{-1}(1 - \epsilon) < b$, for each $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$ and since $g_{nh,\beta}$ stays away from zero with probability tending to one if $\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon$ we get

$$\begin{aligned}
& \int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 dG(t, x) \\
& \lesssim \int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 dG(t, x)
\end{aligned}$$

Furthermore

$$\begin{aligned}
\mathbb{E} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 &= \mathbb{E} \left\{ \int \delta K_h(t - \beta x - u + \beta y) d(\mathbb{P}_n - P_0)(u, y, \delta) \right\}^2 \\
&= O\left(\frac{1}{nh}\right),
\end{aligned}$$

uniformly for (t, x) in a bounded region, so we get

$$\mathbb{E} \int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 dG(t, x) = O\left(\frac{1}{nh}\right).$$

Hence

$$\int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh}\right).$$

The second term on the right-hand side of (6.3) can be treated in a similar way. So we get (3.4). This proves part (i).

We next replace dG in part (i) by $d\mathbb{G}_n$ and we get

$$\begin{aligned}
& \int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 d\mathbb{G}_n(t, x) \\
& \lesssim \int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 d\mathbb{G}_n(t, x) \\
& = \frac{1}{n} \sum_{i=1}^n \{g_{nh,1,\beta}(T_i - \beta X_i) - \mathbb{E}g_{nh,1,\beta}(T_i - \beta X_i)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_i - \beta X_i) < 1 - \epsilon\}}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \mathbb{E} \frac{1}{n} \sum_{i=1}^n \{g_{nh,1,\beta}(T_i - \beta X_i) - \mathbb{E}g_{nh,1,\beta}(T_i - \beta X_i)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_i - \beta X_i) < 1 - \epsilon\}} \\
&= \mathbb{E} \{g_{nh,1,\beta}(T_1 - \beta X_1) - \mathbb{E}g_{nh,1,\beta}(T_1 - \beta X_1)\}^2 1_{\{\epsilon < F_{nh,\beta}(T_1 - \beta X_1) < 1 - \epsilon\}} \\
&\lesssim \mathbb{E} \int_{\epsilon/2 < F_{\beta}(t - \beta x) < 1 - \epsilon/2} \{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)\}^2 dG(t, x) \\
&= O\left(\frac{1}{nh}\right).
\end{aligned}$$

This implies by the Markov inequality,

$$\int_{\epsilon < F_{nh,\beta}(t - \beta x) < 1 - \epsilon} \left\{ \frac{g_{nh,1,\beta}(t - \beta x) - \mathbb{E}g_{nh,1,\beta}(t - \beta x)}{g_{nh,\beta}(t - \beta x)} \right\}^2 d\mathbb{G}_n(t, x) = O_p\left(\frac{1}{nh}\right).$$

The other term on the right-hand side of (6.3) is treated similarly and the first part of (iii) follows.

Part(ii)

We have:

$$(6.4) \quad \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) = \frac{\int (y - x) \{\delta - F_{nh,\beta}(t - \beta x)\} K'_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(t - \beta x)}.$$

We consider the numerator of (6.4). It can be rewritten as

$$\begin{aligned}
& \int (y - x) \{\delta - F_0(u - \beta_0 y)\} K'_h(t - \beta x - u + \beta y) d\mathbb{P}_n(u, y, \delta) \\
&+ \int (y - x) \{F_0(u - \beta_0 y) - F_{\beta}(t - \beta x)\} K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\
&+ \{F_{\beta}(t - \beta x) - F_{nh,\beta}(t - \beta x)\} \int (y - x) K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y).
\end{aligned}$$

The first term can be written as

$$A_n(t, x, \beta) \stackrel{\text{def}}{=} \int (y - x) \{\delta - F_0(u - \beta_0 y)\} K'_h(t - \beta x - u + \beta y) d(\mathbb{P}_n - P_0)(u, y, \delta),$$

and we have:

$$\begin{aligned}
& \mathbb{E} \int_{F_{nh,\beta}(t - \beta x) \in (\epsilon, 1 - \epsilon)} A_n(t, x, \beta)^2 dG(t, x) \leq \mathbb{E} \int A_n(t, x, \beta)^2 dG(t, x) \\
&\sim \frac{1}{nh^3} \int \text{var}(X|v) F_0(v) \{1 - F_0(v)\} f_{T - \beta X}(v) dv \int K'(u)^2 du, \quad n \rightarrow \infty.
\end{aligned}$$

In the second term we must compare $F_0(u - \beta_0 y)$ with

$$F_{\beta}(t - \beta x) = \int F_0(t - \beta_0 x + (\beta - \beta_0)(z - x)) f_{X|T - \beta X}(z|t - \beta x) dz.$$

We can write

$$\begin{aligned} & F_0(u - \beta_0 y) - F_\beta(t - \beta x) \\ &= \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T-\beta X}(z|t - \beta x) dz. \end{aligned}$$

So we find for the second term

$$\begin{aligned} B_n(t, x, \beta) &\stackrel{\text{def}}{=} \int (y - x) \{F_0(u - \beta_0 y) - F_\beta(t - \beta x)\} K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\ &= \int \int (y - x) \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T-\beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\ &= \int (y - x) \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T-\beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) dG(u, y) \\ &\quad + \int (y - x) \int \{F_0(u - \beta_0 y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T-\beta X}(z|t - \beta x) dz \\ &\quad \cdot K'_h(t - \beta x - u + \beta y) d(\mathbb{G}_n - G)(u, y). \end{aligned}$$

We now get for the first term on the right-hand side

$$\begin{aligned} & \int (y - x) f_0(v + (\beta - \beta_0)y) K_h(t - \beta x - v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dv dy \\ & \quad + \int (y - x) \int \{F_0(v + (\beta - \beta_0)y) - F_0(t - \beta_0 x + (\beta - \beta_0)(z - x))\} f_{X|T-\beta X}(z|t - \beta x) dz \\ & \quad \cdot K_h(t - \beta x - v) \frac{\partial}{\partial v} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|v)\} dv dy \\ &= \int (y - x) f_0(v + (\beta - \beta_0)y) K_h(t - \beta x - v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dv dy \\ & \quad + \int (y - x) \int \{F_0(v + (\beta - \beta_0)y) - F_\beta(t - \beta x)\} \\ & \quad \cdot K_h(t - \beta x - v) \frac{\partial}{\partial v} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|v)\} dv dy \\ &= f_{T-\beta X}(t - \beta x) \int (y - x) f_0(t - \beta x + (\beta - \beta_0)y) f_{X|T-\beta X}(y|t - \beta x) dy \\ & \quad + \int (y - x) \{F_0(t - \beta x + (\beta - \beta_0)y) - F_\beta(t - \beta x)\} \\ & \quad \cdot \frac{\partial}{\partial v} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|v)\} \Big|_{v=t-\beta x} dy + O(h) \\ &= f_{T-\beta X}(t - \beta x) a_\beta(t - \beta x) + O(h). \end{aligned}$$

Since

$$g_{nh,\beta}(t - \beta x) = f_{T-\beta X}(t - \beta x) + O_p(h^2),$$

we get,

$$\int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ \frac{B_n(t, x, \beta)}{g_{nh,\beta}(t - \beta x)} - a_\beta(t - \beta x) \right\}^2 dG(t, x) = O_p\left(\frac{1}{nh^3}\right) + O_p(h^2).$$

Finally, defining

$$C_n(t, x, \beta) \stackrel{\text{def}}{=} \{F_\beta(t - \beta x) - F_{nh, \beta}(t - \beta x)\} \int (y - x) K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y),$$

we get, using,

$$\begin{aligned} & \int (y - x) K'_h(t - \beta x - u + \beta y) d\mathbb{G}_n(u, y) \\ &= \int (y - x) K'_h(t - \beta x - u + \beta y) dG(u, y) + \int (y - x) K'_h(t - \beta x - u + \beta y) d(\mathbb{G}_n - G)(u, y) \\ &= \int (y - x) K'_h(t - \beta x - v) f_{T-\beta X}(v) f_{X|T-\beta X}(y|v) dv dy + O_p\left(\frac{1}{nh^3}\right) \\ &= \int (y - x) K_h(t - \beta x - v) \frac{d}{dv} \{f_{T-\beta X}(v) f_{X|T-\beta X}(y|v)\} dv dy + O_p\left(\frac{1}{nh^3}\right) \\ &= O_p(1), \end{aligned}$$

and using Lemma 3.1 for the factor $F_\beta(t - \beta x) - F_{nh, \beta}(t - \beta x)$ that

$$\int_{F_{nh, \beta}(t - \beta x) \in (\epsilon, 1 - \epsilon)} C_n(t, x, \beta)^2 dG(t, x) = O_p\left(\frac{1}{nh}\right) + O_p(h^4).$$

This proves part (ii). The second part of (iii) is proved in the same way as the first part of (iii). \square

PROOF OF LEMMA 3.2. The function $F_{nh, \beta}$ is the ratio of two kernel estimators. If $h \asymp n^{-1/5}$, the derivative has the property

$$\int_{\epsilon/2 < F_\beta(u) < 1 - \epsilon/2} \{F'_{nh, \beta}(u) - F'_\beta(u)\}^2 du = O_p(1),$$

using Proposition 5.1.9, p. 393 in [8], with $m = 1$, $p = 2$ and $h \asymp n^{-1/5}$. So we may assume that $F_{nh, \beta}$ belongs to a class of functions \mathcal{F} with the property that

$$\int_{\epsilon/2 < F_\beta(u) < 1 - \epsilon/2} f'(u)^2 du \leq M.$$

if $f \in \mathcal{F}$, for a fixed $M > 0$. It now follows from [1] that the entropy for the supremum norm H_∞ satisfies:

$$H_\infty(\zeta, \mathcal{F}) = O(\zeta^{-1}),$$

see also the discussion in [27] in section 2.4, since these functions (seen as functions of $t - \beta x$) belong to \mathcal{F} with probability tending to one. We can deduce from this that, for fixed β , the functions in the class \mathcal{F}' :

$$(t, x) \mapsto F_{nh, \beta}(t - \beta x)$$

also have entropy $H_\infty(\zeta, \mathcal{F}') = O(\zeta^{-1})$, and since

$$H_{2, B}(\zeta, \mathcal{F}', G) \leq H_\infty(\zeta, \mathcal{F}')$$

for the entropy with bracketing for the $L_2(G)$ metric, we also have

$$H_{2, B}(\zeta, \mathcal{F}', G) = O(\zeta^{-1}).$$

The function δ can be represented in the form $1_{[e,\infty)}(t - \beta x)$, where e denotes a realization of the errors ε_i , and it is clear that this class of functions, for varying β , also has an ζ -entropy with bracketing for the $L_2(Q_0)$ metric bounded by a constant times ζ^{-1} , where Q_0 is the probability measure of (T, X, ε) . Finally the functions

$$(t, x, e, \beta) \mapsto a_\beta(t - \beta x) \frac{1_{[e,\infty)}(t - \beta x) - F_{nh,\beta}(t - \beta x)}{F_{nh,\beta}(t - \beta x)\{1 - F_{nh,\beta}(t - \beta x)\}},$$

where we let also the parameter β vary, clearly still have the ζ -entropy with bracketing for the $L_2(Q_0)$ metric of order ζ^{-1} .

Applying the equicontinuity lemma to these functions, using the envelope function

$$(t, x, \varepsilon, \beta) \mapsto \sup_{\beta \in [\beta_0 - \eta, \beta_0 + \eta]} \frac{8|a_\beta(t - \beta x)|}{\varepsilon(1 - \varepsilon)},$$

and the seminorm

$$\rho_{Q_0}(f_{1,\beta}, f_{2,\beta'}) = \left\{ \int_{\varepsilon/2 < F_0(t - \beta_0 x) < 1 - \varepsilon/2} \{f_1(t, x, e, \beta) - f_2(t, x, e, \beta')\}^2 dQ_0(t, x, e) \right\}^{1/2}$$

combined with the first result of Lemma 3.1 yields that the limit distribution of

$$\sqrt{n} \int_{F_{nh,\beta}(t - \beta x) \in (\varepsilon, 1 - \varepsilon)} a_\beta(t - \beta x) \frac{\delta - F_{nh,\beta}(t - \beta x)}{F_{nh,\beta}(t - \beta x)\{1 - F_{nh,\beta}(t - \beta x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta)$$

is the same as the limit distribution of

$$\sqrt{n} \int_{F_\beta(t - \beta x) \in (\varepsilon, 1 - \varepsilon)} a_\beta(t - \beta x) \frac{\delta - F_\beta(t - \beta x)}{F_\beta(t - \beta x)\{1 - F_\beta(t - \beta x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta)$$

where we assume that $\{\varepsilon < F_\beta(t - \beta x) < 1 - \varepsilon\}$ is contained in $\{\varepsilon/2 < F_0(t - \beta_0 x) < 1 - \varepsilon/2\}$, for all $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$. This completes the proof of Lemma 3.2. \square

PROOF OF LEMMA 3.3. Let $\beta \in [\beta_0 - \eta, \beta_0 + \eta]$. We have:

$$\begin{aligned} & \sqrt{n} \int_{F_{nh,\beta}(t - \beta x) \in (\varepsilon, 1 - \varepsilon)} \left\{ a_\beta(t - \beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) \right\} \\ & \quad \cdot \frac{\delta - F_{nh,\beta}(t - \beta x)}{F_{nh,\beta}(t - \beta x)\{1 - F_{nh,\beta}(t - \beta x)\}} d\mathbb{P}_n(t, x, \delta) \\ & = \sqrt{n} \int_{F_{nh,\beta}(t - \beta x) \in (\varepsilon, 1 - \varepsilon)} \left\{ a_\beta(t - \beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) \right\} \\ & \quad \cdot \left\{ \frac{\delta - F_0(t - \beta_0 x)}{F_{nh,\beta}(t - \beta x)\{1 - F_{nh,\beta}(t - \beta x)\}} \right\} d\mathbb{P}_n(t, x, \delta) \\ & \quad + \sqrt{n} \int_{F_{nh,\beta}(t - \beta x) \in (\varepsilon, 1 - \varepsilon)} \left\{ a_\beta(t - \beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t - \beta x) \right\} \\ & \quad \cdot \left\{ \frac{F_0(t - \beta_0 x) - F_{nh,\beta}(t - \beta x)}{F_{nh,\beta}(t - \beta x)\{1 - F_{nh,\beta}(t - \beta x)\}} \right\} d\mathbb{G}_n(t, x). \end{aligned}$$

The first term on the right can be written

$$\sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ a_\beta(t-\beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t-\beta x) \Big|_{\beta=\beta} \right\} \cdot \left\{ \frac{\delta - F_0(t-\beta_0 x)}{F_{nh,\beta}(t-\beta x) \{1 - F_{nh,\beta}(t-\beta x)\}} \right\} d(\mathbb{P}_n - P_0)(t, x, \delta),$$

Now note that

$$\frac{\partial}{\partial \beta} F_{nh,\beta}(t-\beta x) = \frac{\int (y-x) \{\delta - F_{nh,\beta}(t-\beta x)\} K'_h(v-u+\beta y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(t-\beta x)}$$

and that the function

$$v \mapsto \frac{\int (y-x) \{\delta - F_{nh,\beta}(v)\} K'_h(v-u+\beta y) d\mathbb{P}_n(u, y, \delta)}{g_{nh,\beta}(v)}$$

has a derivative which is square integrable on the set $\{v : \epsilon/2 < F_\beta(v) < 1 - \epsilon/2\}$, with probability tending to one. So we may assume that the function, as a function of $t - \beta x$, belongs to a class of functions \mathcal{F} with the property that

$$\int_{\epsilon/2 < F_\beta(u) < 1-\epsilon/2} f'(u)^2 du \leq M.$$

if $f \in \mathcal{F}$, for a fixed $M > 0$ (see also the proof of Lemma 3.1; we use Proposition 5.1.9, p. 393 in [8], with $m = 1$, $p = 2$), and $h \asymp n^{-1/5}$).

So we can apply the equicontinuity lemma (see [20], p. 151) which tells us that

$$\begin{aligned} & \sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ a_\beta(t-\beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t-\beta x) \right\} \cdot \left\{ \frac{\delta - F_0(t-\beta_0 x)}{F_{nh,\beta}(t-\beta x) \{1 - F_{nh,\beta}(t-\beta x)\}} \right\} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &= o_p(1), \end{aligned}$$

(see again the proof of Lemma 3.1).

Furthermore, an application of the Cauchy-Schwarz inequality and Lemma 3.1 yield that

$$\begin{aligned} & \sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ a_\beta(t-\beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t-\beta x) \right\} \cdot \left\{ \frac{F_0(t-\beta_0 x) - F_{nh,\beta}(t-\beta x)}{F_{nh,\beta}(t-\beta x) \{1 - F_{nh,\beta}(t-\beta x)\}} \right\} d\mathbb{G}_n(t, x) \\ &= O_p\left(n^{-1/10}\right) + o_p\left(\sqrt{n}(\beta - \beta_0)\right), \quad \beta \rightarrow \beta_0. \end{aligned}$$

The conclusion is that

$$\begin{aligned} & \sqrt{n} \int_{F_{nh,\beta}(t-\beta x) \in (\epsilon, 1-\epsilon)} \left\{ a_\beta(t-\beta x) - \frac{\partial}{\partial \beta} F_{nh,\beta}(t-\beta x) \right\} \cdot \frac{\delta - F_{nh,\beta}(t-\beta x)}{F_{nh,\beta}(t-\beta x) \{1 - F_{nh,\beta}(t-\beta x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= o_p\left(1 + \sqrt{n}(\beta - \beta_0)\right), \quad \beta \rightarrow \beta_0, \end{aligned}$$

and hence that

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1 - \epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{\delta - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= o_p\left(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)\right), \end{aligned}$$

since

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1 - \epsilon)} \frac{\partial}{\partial \beta} F_{nh, \beta}(t - \beta x) \Big|_{\beta = \hat{\beta}_n} \frac{\delta - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} d\mathbb{P}_n(t, x, \delta) \\ &= 0. \end{aligned}$$

We now find

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1 - \epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{\delta - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} dP_0(t, x, \delta) \\ &= \sqrt{n} \int_{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \in (\epsilon, 1 - \epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{F_0(t - \beta_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)}{F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x) \{1 - F_{nh, \hat{\beta}_n}(t - \hat{\beta}_n x)\}} dG(t, x) \\ &\sim \sqrt{n} \int_{F_0(t - \beta_0 x) \in (\epsilon, 1 - \epsilon)} a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \frac{F_0(t - \beta_0 x) - F_{\hat{\beta}_n}(t - \hat{\beta}_n x)}{F_0(t - \beta_0 x) \{1 - F_0(t - \beta_0 x)\}} dG(t, x) \\ &\sim -\sqrt{n}(\hat{\beta}_n - \beta_0) I_\epsilon(\beta_0), \end{aligned}$$

where the last equality follows from an expansion of $a_{\hat{\beta}_n}(t - \hat{\beta}_n x) \{F_0(t - \beta_0 x) - F_{\hat{\beta}_n}(t - \hat{\beta}_n x)\}$. The result now follows. \square

PROOF OF THEOREM 3.3. Using the equicontinuity lemma, we have for all sequences (β_n) such that $\beta_n \rightarrow \beta_0$:

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \beta}(t - \beta_n x) \in (\epsilon, 1 - \epsilon)} a_{\beta_n}(t - \beta_n x) \frac{\delta - F_{nh, \beta_n}(t - \beta_n x)}{F_{nh, \beta_n}(t - \beta_n x) \{1 - F_{nh, \beta_n}(t - \beta_n x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &= \sqrt{n} \int_{F_{nh, \beta_0}(t - \beta_0 x) \in (\epsilon, 1 - \epsilon)} a_{\beta_0}(t - \beta_0 x) \frac{\delta - F_{nh, \beta_0}(t - \beta_0 x)}{F_{nh, \beta_0}(t - \beta_0 x) \{1 - F_{nh, \beta_0}(t - \beta_0 x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &\quad + O_p(\beta_n - \beta_0) \\ &= \sqrt{n} \int_{F_{nh, \beta_0}(t - \beta_0 x) \in (\epsilon, 1 - \epsilon)} a_{\beta_0}(t - \beta_0 x) \frac{\delta - F_0(t - \beta_0 x)}{F_0(t - \beta_0 x) \{1 - F_0(t - \beta_0 x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &\quad + o_p(1), \end{aligned}$$

where

$$a_{\beta_0}(t - \beta_0 x) = f_0(t - \beta_0 x) E_{\beta_0} \{X - x | T - \beta_0 X = t - \beta_0 x\}.$$

Since

$$\sqrt{n} \int_{F_{nh, \beta_0}(t - \beta_0 x) \in (\epsilon, 1 - \epsilon)} a_{\beta_0}(t - \beta_0 x) \frac{\delta - F_0(t - \beta_0 x)}{F_0(t - \beta_0 x) \{1 - F_0(t - \beta_0 x)\}} dP_0(t, x, \delta) = 0,$$

we get:

$$\begin{aligned} & \sqrt{n} \int_{F_{nh, \beta_0}(t - \beta_0 x) \in (\epsilon, 1 - \epsilon)} a_{\beta_0}(t - \beta_0 x) \frac{\delta - F_0(t - \beta_0 x)}{F_0(t - \beta_0 x) \{1 - F_0(t - \beta_0 x)\}} d(\mathbb{P}_n - P_0)(t, x, \delta) \\ &= n^{-1/2} \sum_{i=1}^n f_0(T_i - \beta_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta_0 X_i) - X_i\} \\ & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta_0 X_i)}{F_0(T_i - \beta_0 X_i) \{1 - F_0(T_i - \beta_0 X_i)\}} 1_{(\epsilon, 1 - \epsilon)} \{F_0(T_i - \beta_0 X_i)\}. \end{aligned}$$

Hence, using Lemma 3.3, we get

$$\begin{aligned} & \sqrt{n} I_\epsilon(\beta_0)(\hat{\beta}_n - \beta_0) \\ &= n^{-1/2} \sum_{i=1}^n f_0(T_i - \beta_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta_0 X_i) - X_i\} \\ & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta_0 X_i)}{F_0(T_i - \beta_0 X_i) \{1 - F_0(T_i - \beta_0 X_i)\}} 1_{(\epsilon, 1 - \epsilon)} \{F_0(T_i - \beta_0 X_i)\} + o_p(1). \end{aligned}$$

□

PROOF OF THEOREM 4.1. We will denote $dx_1 \dots dx_k$ by dx . We have

$$\begin{aligned} \hat{\alpha}_n - \alpha_0 &= \int u dF_{nh, \hat{\beta}_n}(u) - \int u dF_0(u) = \int \{F_0(u) - F_{nh, \hat{\beta}_n}(u)\} du \\ &= \int \frac{F_0(t - \hat{\beta}'_n x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\ (6.5) \quad &= \int \frac{F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) + \int \frac{F_0(t - \beta'_0 x) - F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \end{aligned}$$

For the first term in the last expression we get

$$\begin{aligned} & \int \frac{F_0(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x)}{f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\ &= \int \{F_0(u) - F_0(u + x'(\hat{\beta}_n - \beta_0))\} f_{X|T - \hat{\beta}'_n X}(x | T - \hat{\beta}'_n X = u) du dx \\ &\sim - \int x'(\hat{\beta}_n - \beta_0) f_0(u) f_{X|T - \beta'_0 X}(x | T - \beta'_0 X = u) du dx \\ &\sim - \left\{ \int E_{\beta_0} \{X' | T - \beta'_0 X = u\} f_0(u) du \right\} (\hat{\beta}_n - \beta_0) \end{aligned}$$

This term, multiplied with \sqrt{n} , is asymptotically normal, with expectation zero and variance

$$\sigma_1^2 \stackrel{\text{def}}{=} a(\beta_0)' I_\epsilon(\beta_0)^{-1} a(\beta_0),$$

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

$$a(\beta_0) = \int E_{\beta_0} \{X | T - \beta'_0 X = u\} f_0(u) du.$$

For the second term in (6.5), we first note that,

$$(6.6) \quad F_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) - F_0(t - \beta'_0 x) = \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) d\mathbb{P}_n(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x)}.$$

We write (6.6) as the sum of the integral over dP_0 and the integral over $d(\mathbb{P}_n - P_0)$ and show that the contribution of the dP_0 integral, evaluated in (6.5) is negligible and that the contribution of the $d(\mathbb{P}_n - P_0)$ integral will yield an asymptotic normal distribution.

We have

$$\begin{aligned} & \int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dP_0(u, y, \delta) \\ &= \int \{F_0(u - \beta'_0 y) - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dG(u, y) \\ &= \int \{F_0(v + (\hat{\beta}_n - \beta_0)y) - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - v) \\ & \quad \cdot f_{T - \hat{\beta}'_n X}(v) f_{X|T - \hat{\beta}'_n X}(y|T - \hat{\beta}'_n X = v) dv dy \\ &= f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x) \int \{F_0(t - \hat{\beta}'_n x + (\hat{\beta}_n - \beta_0)y) - F_0(t - \beta'_0 x)\} \\ & \quad \cdot f_{X|T - \hat{\beta}'_n X}(y|T - \hat{\beta}'_n X = t - \hat{\beta}'_n x) dy + O_p(h^2) \\ &= f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x) f_0(t - \beta'_0 x) (\hat{\beta}_n - \beta_0)' E\{X - x|T - \hat{\beta}'_n X = t - \hat{\beta}'_n x\} \\ & \quad + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|), \end{aligned}$$

where $\|x\|$ is the euclidean norm of the vector x . Hence we get

$$\begin{aligned} & \int \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) dP_0(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\ &= (\hat{\beta}_n - \beta_0)' \int \frac{f_0(t - \beta'_0 x) E\{X - x|T - \beta_0 X = t - \beta'_0 x\}}{g_{nh}(t - \hat{\beta}'_n x)} dG(t, x) + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|) \\ &= (\hat{\beta}_n - \beta_0)' \int f_0(v) E\{X - x|T - \beta_0 X = v\} f_{X|T - \beta_0 X}(x|T - \beta_0 X = v) dx dv \\ & \quad + O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|) \\ &= O_p(h^2) + o_p(\|\hat{\beta}_n - \beta_0\|), \end{aligned}$$

which is $o_p(n^{-1/2})$ if $h \ll n^{-1/4}$.

Finally,

$$\begin{aligned} & \sqrt{n} \int \frac{\int \{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y) d(\mathbb{P}_n - P_0)(u, y, \delta)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) \\ &= \sqrt{n} \iint \frac{\{\delta - F_0(t - \beta'_0 x)\} K_h(t - \hat{\beta}'_n x - u + \hat{\beta}'_n y)}{g_{nh, \hat{\beta}_n}(t - \hat{\beta}'_n x) f_{T - \hat{\beta}'_n X}(t - \hat{\beta}'_n x)} dG(t, x) d(\mathbb{P}_n - P_0)(u, y, \delta) \\ &= \sqrt{n} \int \frac{\{\delta - F_0(u - \beta'_0 y)\}}{f_{T - \beta_0 X}(u - \beta'_0 y)} d(\mathbb{P}_n - P_0)(u, y, \delta) + O_p(h^2) + O_p(\|\hat{\beta}_n - \beta_0\|) \end{aligned}$$

is asymptotically normal, with expectation zero and variance

$$(6.7) \quad \int \frac{F_0(v)\{1 - F_0(v)\}}{f_{T-\beta_0 X}(v)} dv,$$

if $h \ll n^{-1/4}$.

Both terms in the representation on the right of (6.5) are, apart from a negligible contribution, sums of independent variables with expectation zero. By Theorem 3.3 we have

$$\begin{aligned} & \sqrt{n}(\hat{\beta}_n - \beta_0) \\ &= n^{-1/2} I_\epsilon(\beta_0)^{-1} \sum_{i=1}^n f_0(T_i - \beta_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta'_0 X_i) - X_i\} \\ & \quad \cdot \frac{\Delta_i - F_0(T_i - \beta_0 X_i)}{F_0(T_i - \beta_0 X_i) \{1 - F_0(T_i - \beta_0 X_i)\}} 1_{(\epsilon, 1-\epsilon)} \{F_0(T_i - \beta_0 X_i)\} + o_p(1). \end{aligned}$$

and the second term of (6.5) has the representation

$$n^{-1/2} \sum_{i=1}^n \frac{\Delta_i - F_0(T_i - \beta'_0 X_i)}{f_{T-\beta'_0 X}(T_i - \beta'_0 X_i)}.$$

By the independence of the summands with indices $i \neq j$, the only contribution to the covariance of the two terms in the representation can come from summands with the same index. But,

$$\begin{aligned} & E_{\beta_0} \left\{ \frac{f_0(T_i - \beta'_0 X_i) \{E_{\beta_0}(X_i | T_i - \beta'_0 X_i) - X_i\} \{\Delta_i - F_0(T_i - \beta'_0 X_i)\}^2}{F_0(T_i - \beta'_0 X_i) \{1 - F_0(T_i - \beta'_0 X_i)\} f_{T-\beta'_0 X}(T_i - \beta'_0 X_i)} 1_{(\epsilon, 1-\epsilon)} \{F_0(T_i - \beta'_0 X_i)\} \right\}. \\ &= \int_{F_0(u - \beta'_0 y) \in (\epsilon, 1-\epsilon)} \frac{f_0(u - \beta'_0 y) \{E_{\beta_0}(X | T - \beta'_0 X = u - \beta'_0 y) - y\} \{\delta - F_0(u - \beta'_0 y)\}^2}{F_0(u - \beta'_0 y) \{1 - F_0(u - \beta'_0 y)\} f_{T-\beta'_0 X}(u - \beta'_0 y)} dP_0(u, y, \delta) \\ &= \iint_{F_0(v) \in (\epsilon, 1-\epsilon)} \frac{f_0(v) \{E_{\beta_0}(X | T - \beta'_0 X = v) - y\} F_0(v) \{1 - F_0(v)\}}{F_0(v) \{1 - F_0(v)\}} f_{X|T-\beta'_0 X}(y|v) dv dy \\ &= \int_{F_0(v) \in (\epsilon, 1-\epsilon)} \int \{E_{\beta_0}(X | T - \beta'_0 X = v) - y\} f_{X|T-\beta'_0 X}(y|v) dy \frac{f_0(v) F_0(v) \{1 - F_0(v)\}}{F_0(v) \{1 - F_0(v)\}} dv \\ &= 0 \end{aligned}$$

So the covariance is zero and Theorem 4.1 follows. \square

Finally, for completeness, we give the representation of the partial derivatives of the constants c_1 and c_2 w.r.t. β , which was used in computing the derivative $\psi_n^{(\epsilon)}(\beta, \bar{F}_{nh, \beta})$ in Figure 4. We have

$$\begin{aligned} & \frac{\partial}{\partial \beta} c_1 \\ &= - \frac{\int_{\bar{a} < u - \beta' y < \bar{b}} y e^{-(u - \beta' y - \bar{a})/\sqrt{\lambda}} \hat{F}_{n, \beta}(u - \beta' y) / f_{T-\beta' X}(u - \beta' y) d\mathbb{G}_n(u, y)}{2\lambda \left\{ 1 - e^{-2(\bar{b} - \bar{a})/\sqrt{\lambda}} \right\}} \\ & \quad - \frac{e^{-(\bar{b} - \bar{a})/\sqrt{\lambda}} \int_{\bar{a} < u - \beta' y < \bar{b}} y e^{-(\bar{b} - u + \beta' y)/\sqrt{\lambda}} \hat{F}_{n, \beta}(u - \beta' y) / f_{T-\beta X}(u - \beta' y) d\mathbb{G}_n(u, y)}{2\lambda \left\{ 1 - e^{-2(\bar{b} - \bar{a})/\sqrt{\lambda}} \right\}} \\ & \quad - \frac{\int_{\bar{a} < u - \beta' y < \bar{b}} e^{-(u - \beta' y - \bar{a})/\sqrt{\lambda}} y \hat{F}_{n, \beta}(u - \beta' y) f'_{T-\beta X}(u - \beta' y) / f_{T-\beta' X}(u - \beta' y)^2 d\mathbb{G}_n(u, y)}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b} - \bar{a})/\sqrt{\lambda}} \right\}} \end{aligned}$$

$$\frac{e^{-(\bar{b}-\bar{a})/\sqrt{\lambda}} \int_{\bar{a}<u-\beta'y<\bar{b}} y e^{-(\bar{b}-u+\beta'y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(u-\beta'y) f'_{T-\beta X}(u-\beta'y) / f_{T-\beta X}(u-\beta'y)^2 d\mathbb{G}_n(u, y)}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}},$$

and

$$\begin{aligned} & \frac{\partial}{\partial \beta} c_2 \\ &= \frac{e^{-(\bar{b}-\bar{a})/\sqrt{\lambda}} \int_{\bar{a}<u-\beta'y<\bar{b}} y e^{-(u-\beta'y-\bar{a})/\sqrt{\lambda}} \hat{F}_{n,\beta}(u-\beta'y) / f_{T-\beta X}(u-\beta'y) d\mathbb{G}_n(u, y)}{2\lambda \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}} \\ &+ \frac{\int_{\bar{a}<u-\beta'y<\bar{b}} y e^{-(\bar{b}-u+\beta'y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(u-\beta'y) / f_{T-\beta X}(u-\beta'y) d\mathbb{G}_n(u, y)}{2\lambda \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}} \\ &+ \frac{e^{-(\bar{b}-\bar{a})/\sqrt{\lambda}} \int_{\bar{a}<u-\beta'y<\bar{b}} y e^{-(u-\beta'y-\bar{a})/\sqrt{\lambda}} \hat{F}_{n,\beta}(u-\beta'y) f'_{T-\beta X}(u-\beta'y) / f_{T-\beta X}(u-\beta'y)^2 d\mathbb{G}_n(u, y)}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}} \\ &+ \frac{\int_{\bar{a}<u-\beta'y<\bar{b}} y e^{-(\bar{b}-u+\beta'y)/\sqrt{\lambda}} \hat{F}_{n,\beta}(u-\beta'y) f'_{T-\beta X}(u-\beta'y) / f_{T-\beta X}(u-\beta'y)^2 d\mathbb{G}_n(u, y)}{2\sqrt{\lambda} \left\{ 1 - e^{-2(\bar{b}-\bar{a})/\sqrt{\lambda}} \right\}}. \end{aligned}$$

7. Acknowledgements. The authors want to thank Richard Nickl for useful comments and his reference to relevant theory in the book [8] (to appear). The research of the second author was supported by the Research Foundation Flanders (FWO) [grant number 11W7315N]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged.

References.

- [1] BIRMAN, M. Š. and SOLOMJAK, M. Z. (1967). Piecewise polynomial approximations of functions of classes W_p^α . *Mat. Sb. (N.S.)* **73 (115)** 331–355. [MR0217487 \(36 ##576\)](#)
- [2] COSSLETT, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51** 765–782. [MR712369 \(85a:62174\)](#)
- [3] COSSLETT, S. R. (1987). Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* **55** 559–585. [MR890854 \(88e:62087\)](#)
- [4] COSSLETT, S. R. (2007). Efficient estimation of semiparametric models by smoothed maximum likelihood. *Internat. Econom. Rev.* **48** 1245–1272. [MR2375624](#)
- [5] 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 \(2006h:60039\)](#)
- [6] FINKELSTEIN, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* **42** 845–854. [MR872963 \(87m:62286\)](#)
- [7] FINKELSTEIN, D. M. and WOLFE, R. A. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945.
- [8] GINÉ, E. and NICKL, R. (2015). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. To appear in Cambridge Series in Statistical and Probabilistic Mathematics.
- [9] 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.
- [10] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge.
- [11] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. *DMV Seminar* **19**. Birkhäuser Verlag, Basel. [MR1180321 \(94k:62056\)](#)
- [12] HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171 \(94d:62134\)](#)

- [13] HUANG, J. (1995). Maximum likelihood estimation for proportional odds regression model with current status data. In *Analysis of censored data (Pune, 1994/1995)*. *IMS Lecture Notes Monogr. Ser.* **27** 129–145. Inst. Math. Statist., Hayward, CA. [MR1483344](#)
- [14] HUANG, J. and WELLNER, J. (1993). Regression models with interval censoring. *Proceedings of the Kolmogorov Seminar, Euler Mathematics Institute. St. Petersburg, Russia*.
- [15] JEWELL, N. P. (1982). Mixtures of exponential distributions. *The Annals of Statistics* 479–484.
- [16] KLEIN, R. W. and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61** 387–421. [MR1209737 \(93k:62214\)](#)
- [17] KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*. *Inst. Math. Stat. Collect.* **1** 282–292. Inst. Math. Statist., Beachwood, OH.
- [18] LI, G. and ZHANG, C.-H. (1998). Linear regression with interval censored data. *Ann. Statist.* **26** 1306–1327. [MR1647661 \(2000b:62094\)](#)
- [19] MURPHY, S. A., VAN DER VAART, A. W. and WELLNER, J. A. (1999). Current status regression. *Math. Methods Statist.* **8** 407–425. [MR1735473 \(2001a:62126\)](#)
- [20] POLLARD, D. (1984). *Convergence of stochastic processes*. *Springer Series in Statistics*. Springer-Verlag, New York. [MR762984 \(86i:60074\)](#)
- [21] RABINOWITZ, D., TSIATIS, A. and ARAGON, J. (1995). Regression with interval censored data. *Biometrika* **82** 501–513.
- [22] ROSSINI, A. and TSIATIS, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *Journal of American Statistical Association* **91** 713–721.
- [23] SEN, B., BANERJEE, M. and WOODROOFE, M. B. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. [MR2676880 \(2011f:62046\)](#)
- [24] SEN, B. and XU, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. [MR3257405](#)
- [25] SHEN, X. (2000). Linear regression with current status data. *J. Amer. Statist. Assoc.* **95** 842–852. [MR1804443 \(2001k:62118\)](#)
- [26] SHIBOSKI, S. C. and JEWELL, N. P. (1992). Statistical Analysis of the Time Dependence of HIV Infectivity Based on Partner Study Data. *Journal of the American Statistical Association* **87** 360–372.
- [27] VAN DE GEER, S. A. (2000). *Applications of empirical process theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge University Press, Cambridge. [MR1739079 \(2001h:62002\)](#)
- [28] VAN DER VAART, A. W. (1998). *Asymptotic statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. [MR1652247 \(2000c:62003\)](#)