Semiparametric Estimation of a Censored Regression Model with an Unknown Transformation of the Dependent Variable

Tue Gørgens and Joel Horowitz*
Department of Economics
The University of Iowa
Iowa City, Iowa 52242

February 28, 1996

Abstract

In this paper we develop semiparametric estimators of Λ and Ψ in the model $\Lambda(Y) = \min[\beta' X + U, C]$, where Y is a nonnegative dependent variable, X is a vector of explanatory variables, U is an unobserved random "error" term with unknown distribution function Ψ, C is a random censoring variable, β is an unknown parameter vector, and Λ is an unknown strictly increasing function. This model includes as a special case the censored proportional hazards model with unobserved heterogeneity. Estimators of Λ and Ψ already exist for the case where either Λ or Ψ belongs to a known finite-dimensional parametric family, and methods for estimating β exist for the general case. In this paper we propose estimators of Λ and Ψ which do not assume that Λ and Ψ belong to known parametric families. We obtain their asymptotic distributions and investigate the small sample properties of the estimators by Monte Carlo simulation.

*We thank Gene Savin, Forrest Nelson and Martin Appel for comments on an earlier draft of this paper. This research was supported in part by NSF grant SBR-9307677.

1 Introduction

A linear uncensored regression model with a transformed dependent variable has the form

$$\Lambda(Y^*) = \beta' X + U, \tag{1}$$

where X is an r-dimensional random vector, U and Y^* are random variables, β is an r-dimensional vector of parameters, and Λ is a strictly increasing real function. In general, β, Λ and the distribution function, Ψ, of U are unknown. Model (1) is closely related to the linear generalized accelerated failure-time model defined by Ridder (1990), $\log f(Y^*) = \beta' X + U$, where $Y^* \geq 0$ and f is a nondecreasing map from $[0, \infty)$ onto $[0, \infty)$.

Model (1) is widely used in applied econometrics. It includes as special cases log-linear models and accelerated failure-time models, transformation models in which Λ is specified up to a vector of parameters as in the Box-Cox transformation, and parametric and semiparametric proportional hazards models (see e.g. Kiefer 1988). An important special case is the mixed proportional hazards model or the proportional hazards model with unobserved heterogeneity (see e.g. Ridder 1990). This model has received a lot of recent attention from researchers trying to separate duration dependence embodied in Λ from unobserved heterogeneity contained in Ψ (see the references below).

Methods for estimating β (up to scale) have been described by e.g. Han (1987), Härdle and Stoker (1989), Horowitz and Härdle (1994), Ichimura (1993) and Powell, Stock, and Stoker (1989). These estimators do not require that either Λ or Ψ be known or belong to known finite-dimensional parametric families of functions. In contrast, most estimators of Λ and Ψ require parametric assumptions about Λ or Ψ (see e.g. Heckman and Singer 1984, Murphy 1991, 1992, and Nielsen, Gill, Andersen, and Sørensen 1992). Recently, however, Horowitz (1996) has shown how to estimate Λ and Ψ without making parametric assumptions. Specifically, Horowitz shows how to obtain $n^{1/2}$-consistent, asymptotically normally distributed nonparametric estimators of Λ and Ψ from a random sample on X and Y^*. To use Horowitz's estimators, β must be estimated beforehand using, for example, one of the estimators proposed in the papers just listed.

In applications, data are often censored; this is particularly true for duration studies. The purpose of this paper is to extend Horowitz's results to the case where some observations are right-censored. The model studied here is

$$\Lambda(Y) = \min[\beta' X + U, C], \quad Y \geq 0, \tag{2}$$

where C is a random censoring variable. Let M be the indicator of no censoring. That is, if $\beta' X + U \leq C$, then Y is not censored and $M = 1$; and if $\beta' X + U > C$, then Y is censored and $M = 0$. The distribution of C may be degenerate and may depend on X, so the assumption of random censoring is not very restrictive.[1] We assume that M, X and Y are observed, but C is unobserved, which is the case in most applications.

This paper focuses on estimating Λ and Ψ in (2) from a random sample of observations on M, X and Y without assuming that Λ and Ψ belong to known finite-dimensional parametric families of functions. Like Horowitz (1996) we assume that a $n^{1/2}$-consistent estimator of β is available. Using empirical process methods we prove uniform consistency and convergence in distribution of the estimators of Λ and Ψ. The method of proof is similar to that used by Horowitz (1996), but the details are more complex owing to the complexity of the estimating equations for Λ and Ψ in the censored case. In addition to proving limit theorems, we investigate the finite-sample properties of the estimators through Monte Carlo experiments.

For the uncensored regression model (1), Horowitz (1996) also shows how to predict Y^* conditional on X using quantiles of the conditional distribution of Y^* given X. He also proves consistency and asymptotic normality of the conditional quantiles of the distribution of Y^*. In the censored model (2) Y^* defined by (1) can be considered a latent, unobserved variable. With some minor changes in the interpretation of the theorem Horowitz's results

[1] Alternatively, with fixed censoring the asymptotic properties of the estimators can be studied conditionally on the observed values of C rather than in terms of the distribution of C.

continue to hold for predicting the latent variable Y^* conditional on X in the censored regression model.

The paper is organized as follows. In section 2 we outline the idea of the estimators. In section 3 we state the theorems establishing their asymptotic behavior. As examples of the small-sample properties of the estimators we report the results of a few Monte Carlo experiments in section 4. Proofs are relegated to appendix A.

Notation: We use $D_i^j f$ to denote the j'th order partial derivative with respect to the i'th argument of the function f, and if f lives on R then $D^j f$ is the j'th order derivative of f. For first order derivatives we often omit the superscript and write $D_i f$ or Df. R^j denotes j-dimensional Euclidean space. The *support* of a function is the closure of the subset of its domain in which it takes nonzero values. The *support* of a random variable or vector is the smallest closed set which contains its values with probability one. If v is a vector, then v_1 denotes the first component and v_{-1} the vector of remaining components of v.

2 The Estimators

Since U and C are unobserved, there is a well-known identification problem. Given the observed variables M, X and Y, equation (2) continues to hold if Λ, U and C are replaced by $\Lambda - \alpha$, $U - \alpha$ and $C - \alpha$ where α is any constant. Equation (2) also holds if Λ, β, U and C are replaced by $\alpha\Lambda$, $\alpha\beta$, αU and αC where α is any positive constant. Therefore, scale and location normalizations are needed. As shown e.g. by Ichimura (1993), identification of β up to scale requires that X has at least one component that is continuously distributed conditional on the others and whose β-coefficient is nonzero. Let this be the first component of X. For scale normalization we set $|\beta_1| = 1$, where β_1 is the first component of β. For location normalization we use $\Lambda(t_0) = 0$, where t_0 is a constant to be determined later.

2.1 Motivation

Let β_n be the given estimator of β, where n is the sample size. The idea is to express Λ and Ψ in terms of the conditional distribution of (Y, M) given $\beta' X$ and then define the estimators Λ_n and Ψ_n by replacing β with β_n and other unknown quantities in these formulae with sample analogs based on kernel estimation.

Define $Z = \beta' X$. Assume that Λ is strictly increasing and differentiable, that U and X are independent, that C and U are conditionally independent given X, and that the distribution function Ψ of U is differentiable. Let $\Theta(z, \cdot)$ be the conditional distribution function of C given $Z = z$. Then the conditional distribution function of Y given $Z = z$ is[2]

$$\Pr(Y \leq y | Z = z) = 1 - [1 - \Psi(\Lambda(y) - z)][1 - \Theta(z, \Lambda(y))]. \tag{3}$$

The conditional sub-distribution function of Y given $Z = z$ when Y is uncensored is given by

$$\Pr(Y \leq y, M = 1 | Z = z) = \int_0^{\Lambda(y) - z} [1 - \Theta(z, z + u)] \, d\Psi(u). \tag{4}$$

Let p_Z denote the density of Z. Define

$$\begin{aligned} A(z, y) &= p_Z(z) \frac{\partial}{\partial y} \Pr(Y \leq y, M = 1 | Z = z) \\ &= p_Z(z) [1 - \Theta(z, \Lambda(y))] D\Psi(\Lambda(y) - z) D\Lambda(y) \end{aligned} \tag{5}$$

and

$$\begin{aligned} B(z, y) &= p_Z(z) \Pr(Y > y | Z = z) \\ &= p_Z(z) [1 - \Theta(z, \Lambda(y))][1 - \Psi(\Lambda(y) - z)]. \end{aligned} \tag{6}$$

Observe that $A(z, y)/B(z, y) = D\Psi(\Lambda(y) - z) D\Lambda(y) \ / \ [1 - \Psi(\Lambda(y) - z)]$. Subject to

[2] Equations (3) and (4) appear for example in Breslow and Crowley (1974).

regularity conditions, some algebra shows that $\int_0^y [A(z,v)/B(z,v)]\,dv = -\log[1-\Psi(\Lambda(y)-z)]$, so $(\partial/\partial z)\int_0^y [A(z,v)/B(z,v)]\,dv = -D\Psi(\Lambda(y)-z)\;/\;[1-\Psi(\Lambda(y)-z)]$. Therefore,

$$\frac{A(z,y)/B(z,y)}{\displaystyle\int_0^y \frac{D_1A(z,v)B(z,v)-A(z,v)D_1B(z,v)}{B(z,v)^2}\,dv} = D\Lambda(y). \tag{7}$$

The proposed estimator of Λ is based on integration of the left-hand side of (7). The integrand in the denominator is simply the partial derivative with respect to z of $A(z,v)/B(z,v)$. Equation (7) holds for all (z,y) such that the denominators are nonzero, the derivatives D_1A and D_1B exist and the integrand is bounded for $0 \le v \le y$.

Let $T_\Lambda \subset R$ be the set where Λ is to be estimated, and normalize Λ by setting $\Lambda(t_0)=0$ for some $t_0 \in T_\Lambda$. Choose a weight function W_Λ on R^3 such that (7) holds for (z,y) in the support $W_\Lambda(\cdot,\cdot,t)$ for all $t \in T_\Lambda$ and such that $W_\Lambda(\cdot,y,t)$ integrates to one whenever $t_0 \le y \le t$ or $t_0 \ge y \ge t$ and $t \in T_\Lambda$. Then for $t \in T_\Lambda$

$$\Lambda(t) = -\iint_{t_0}^{t} W_\Lambda(z,y,t)\left[\frac{A(z,y)/B(z,y)}{\displaystyle\int_0^y \frac{D_1A(z,v)B(z,v)-A(z,v)D_1B(z,v)}{B(z,v)^2}\,dv}\right]dy\,dz, \tag{8}$$

where $\int_{t_0}^{t} = -\int_t^{t_0}$ if $t_0 > t$.[3] The estimator Λ_n is formed by replacing the four unknown functions A, B, D_1A and D_1B in (8) with kernel estimators to be defined shortly. As in Horowitz's estimator for the uncensored case, it is the averaging over z which boosts the slow convergence of the kernel estimators into $n^{1/2}$-consistency of Λ_n.

Once Λ is estimated on T_Λ, Horowitz (1996) uses the empirical distribution of $\Lambda(Y^*)-Z$ to derive an estimator of Ψ, but this is not an option here since Y^* is not observed. Instead, the discussion leading up to (7) shows that when $B(z,y)>0$

$$\Psi(\Lambda(v)-z) = 1-\exp\left(-\int_0^v \frac{A(z,y)}{B(z,y)}\,dy\right). \tag{9}$$

[3]Instead of explicitly restricting the integration to the support of $W_\Lambda(\cdot,\cdot,t)$, we adopt the conventions that $0/0=0$ and $0\cdot$*undefined*$=0$.

The value of Ψ at a given point t is obtained by letting $z = \Lambda(v) - t$ or, equivalently, $v = \Lambda^{-1}(z + t)$. After making the substitution we can average over v or z; as with Λ this improves the rate of convergence of the estimator. Whether we average over v or z does not matter asymptotically, but it affects the estimator in finite samples. In this paper we choose to average over v. Let $T_\Psi \subset R$ be the set where Ψ is to be estimated, and let W_Ψ be a weight function on R^2 such that (9) holds with $z = \Lambda(v) - t$ and $v \in T_\Lambda$ for (v, t) in the support of W_Ψ, and such that $W_\Psi(\cdot, t)$ integrates to 1 for all $t \in T_\Psi$. Then for $t \in T_\Psi$

$$\Psi(t) = 1 - \int W_\Psi(v,t) \exp\Big(- \int_0^v \frac{A(\Lambda(v) - t, y)}{B(\Lambda(v) - t, y)} \, dy \Big) \, dv. \tag{10}$$

The function Ψ is estimated by replacing A, B and Λ in (10) by their estimators. This is the continuous-time equivalent of the Kaplan-Meier estimator of the distribution function of U modified to accomodate the unknown function Λ.

2.2 Implementation

Suppose a random sample $\{(M_i, X_i, Y_i)\}_{i=1}^n$ is observed, and let β_n be a $n^{1/2}$-consistent estimator of β. Let κ_{zn} and κ_{yn} be sequences of positive real numbers converging to 0 as n goes to ∞, the "bandwidths", and let K_z and K_y be real functions on R which integrate to 1, the "kernels", with K_z differentiable. The kernel estimators of A and B are

$$A_n(z, y) = \frac{1}{n} \sum_{i=1}^n M_i \frac{1}{\kappa_{zn}} K_z\Big(\frac{\beta_n' X_i - z}{\kappa_{zn}}\Big) \frac{1}{\kappa_{yn}} K_y\Big(\frac{Y_i - y}{\kappa_{yn}}\Big) \tag{11}$$

and

$$B_n(z, y) = \frac{1}{n} \sum_{i=1}^n 1(Y_i > y) \frac{1}{\kappa_{zn}} K_z\Big(\frac{\beta_n' X_i - z}{\kappa_{zn}}\Big). \tag{12}$$

In general, different kernels and bandwidths could be used in the estimation of the derivatives $D_1 A$ and $D_1 B$, but for simplicity we estimate $D_1 A$ and $D_1 B$ by $D_1 A_n$ and $D_1 B_n$; hence the requirement that K_z be differentiable. The estimator of Λ is defined for $t \in T_\Lambda$

by

$$\Lambda_n(t) = -\iint_{t_0}^{t} W_\Lambda(z,y,t)\left[\frac{A_n(z,y)/B_n(z,y)}{\int_0^y \frac{D_1A_n(z,v)B_n(z,v)-A_n(z,v)D_1B_n(z,v)}{B_n(z,v)^2}\,dv}\right]dy\,dz \tag{13}$$

and the estimator of Ψ for $t \in T_\Psi$ by

$$\Psi_n(t) = 1 - \int W_\Psi(v,t)\exp\Big(-\int_0^v \frac{A_n(\Lambda_n(v)-t,y)}{B_n(\Lambda_n(v)-t,y)}\,dy\Big)\,dv. \tag{14}$$

For definiteness, if an integrand is not defined because of a zero denominator set it equal to zero.

To compute the estimators one must choose T_Λ, W_Λ, T_Ψ, W_Ψ, K_z, K_y, κ_{zn} and κ_{yn}. Precise conditions they must satisfy are given in section 3. Among other things, the kernels must have support in $[-1,1]$ and they must be of "higher order." (A kernel K on R is of order k if $\int K(s)\,ds = 1$ and $\int K(s)s^j\,ds = 0$ for $j = 1,\ldots,k-1$.) Let k_z and k_y be the order of the kernels K_z and K_y, then one possibility is $k_z \geq 5$, $k_y \geq 3$, $\kappa_{zn} \propto n^{-.105}$ and $\kappa_{yn} \propto n^{-.17}$; another is $k_z \geq 7$, $k_y \geq 2$, $\kappa_{zn} \propto n^{-.08}$ and $\kappa_{yn} \propto n^{-.255}$.

We now turn to W_Λ and T_Λ. Let S_Λ denote the union of the supports of $W_\Lambda(\cdot,\cdot,t)$ for all $t \in T_\Lambda$. As mentioned in section 2.1 equation (7) must hold for (z,y) in S_Λ, but this is not sufficient. To begin with, S_Λ must be bounded away from points where A or B are not differentiable, that is, there must be an $\epsilon > 0$ such that the distance is at least ϵ from any point in S_Λ to any point where D_1A or D_1B do not exist. Moreover, the denominators in (8) must be bounded away from zero on S_Λ. In applications one can use the estimates of the denominators to get an idea of the sets where they are larger than some $\epsilon > 0$.[4] Having chosen S_Λ, if the intersection $I_y = \{z : (z,y) \in S_\Lambda\}$ is an interval, one can construct W_Λ

[4] The limit theorems provided in section 3 are derived assuming that the weight functions do not depend on the data.

itself, for example, by

$$W_\Lambda(z,y,t) = f\left(\frac{2z - c_l(y) - c_u(y)}{c_u(y) - c_l(y)}\right)\frac{2}{c_u(y) - c_l(y)}, \tag{15}$$

where $c_l(y)$ and $c_u(y)$ are the lower and upper bounds of I_y and f is a one-dimensional weight function on $[-1,1]$.

The pair T_Ψ and W_Ψ must have properties similar to T_Λ and W_Λ. Let S_Ψ denote the support of W_Ψ. Then y must be in T_Λ if $(z,y) \in S_\Psi$ for some z, and B must be bounded away from zero on the set $\{(\Lambda(y) - t, y) : (y,t) \in S_\Psi\}$. In applications, given $\epsilon > 0$ one can get an idea of the set where $B(\Lambda(y) - t, y) > \epsilon$ from the estimate $B_n(\Lambda_n(y) - t, y)$ and choose S_Ψ accordingly. Given S_Ψ, W_Ψ can be formed as in (15).

We close this section with some remarks on how to compute the estimators. By exploiting the fact that $B_n(z,\cdot)$ is a right-continuous step function jumping down at each Y_i, it is possible to evaluate analytically the inner integrals in the definitions of Λ_n and Ψ_n. Define $a_0 = 0$, $a_k = y$, and let $a_1,\ldots,a_{k-1}$ be the order statistics of the observations Y_i that are less than y. The integral in the denominator in (13) can be computed by

$$\sum_{j=1}^{k} \frac{\left[\sum M_i \int_{a_{j-1}}^{a_j} \frac{1}{\kappa_{yn}} K_y\left(\frac{Y_i - v}{\kappa_{yn}}\right) dv\, K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\right]\left[\sum 1(Y_i > a_j)\frac{1}{\kappa_{zn}} DK_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\right]}{\left[\sum 1(Y_i > a_j) K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\right]^2}$$
$$- \frac{\sum M_i \int_{a_{j-1}}^{a_j} \frac{1}{\kappa_{yn}} K_y\left(\frac{Y_i - v}{\kappa_{yn}}\right) dv\, \frac{1}{\kappa_{zn}} DK_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)}{\sum 1(Y_i > a_j) K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)}, \tag{16}$$

where $\sum$ is short for $\sum_{i=1}^n$. Similarly, define $a_0 = 0$, $a_k = v$, and let $a_1,\ldots,a_{k-1}$ be the order statistics of the observations Y_i that are less than v. Then the inner integral in (14) equals

$$\sum_{j=1}^{k} \frac{\sum M_i \int_{a_{j-1}}^{a_j} \frac{1}{\kappa_{yn}} K_y\left(\frac{Y_i - y}{\kappa_{yn}}\right) dy\, K_z\left(\frac{\beta_n' X_i - \Lambda_n(v) + t}{\kappa_{zn}}\right)}{\sum 1(Y_i > a_j) K_z\left(\frac{\beta_n' X_i - \Lambda_n(v) + t}{\kappa_{zn}}\right)}. \tag{17}$$

For many choices of K_y evaluating $\int_a^b K_y(y)\,dy$ is straightforward; in particular, this is true when K_y is piecewise polynomial. The complexity of the integrands implies that the outer integrals in the definition of Λ_n and Ψ_n must be carried out by numerical methods.

2.3 Predicting Y^* Conditional on X

Using methods similar to those of Horowitz (1996), conditional quantiles of the latent, unobserved variable Y^* can be estimated $n^{1/2}$-consistently. Typically Y^* is predicted using a consistent estimator of $E(Y^*|X=x)$, but this is not an option here because the censoring of Y^* and the need to bound denominators away from zero imply that Λ and Ψ cannot be estimated everywhere on their domains. An alternative is to use an estimator of the median or some other quantile of the conditional distribution of Y^* given $X=x$. For the uncensored model (1), Horowitz shows how to use Λ_n and Ψ_n to obtain estimators of conditional quantiles of Y^*. His derivations carry over to the censored case and are summarized here only because they are needed for the Monte Carlo experiments.

For $\tau \in (0,1)$, let u_τ denote the τ'th quantile of the distribution of U. Since $\Pr(U \leq u_\tau) = \tau$ if and only if $\Pr(Y^* \leq \Lambda^{-1}(\beta' x + u_\tau)|X=x) = \tau$, the τ'th quantile of the distribution of Y^* conditional on $X=x$ is $y_\tau(x) = \Lambda^{-1}(\beta' x + u_\tau)$. The estimator of y_τ is defined as follows. Given u_τ in the interior of T_Ψ, choose a set $T_\tau \subset R^r$ such that $y_\tau(x) \in [\inf T_\Lambda + \epsilon, \sup T_\Lambda - \epsilon]$ for all $x \in T_\tau$, where $\epsilon > 0$ is some constant. Define

$$u_{n\tau} = \inf\{u \in T_\Psi : \Psi_n(u) \geq \tau \text{ or } u = \sup T_\Psi\} \tag{18}$$

and

$$y_{n\tau}(x) = \inf\{y \in T_\Lambda : \Lambda_n(y) \geq \beta_n' x + u_{n\tau} \text{ or } y = \sup T_\Lambda\} \tag{19}$$

Then $y_{\tau n}$ is a uniformly $n^{1/2}$-consistent estimator of y_τ on T_τ.

3 Assumptions and Limit Theorems

Under conditions given in assumptions 1–9 below, the estimators Λ_n and Ψ_n are uniformly $n^{1/2}$-consistent on T_Λ and T_Ψ. Properly centered and normalized they converge weakly to Gaussian stochastic processes with means of zero.

Assumption 1 *The function Λ is real, strictly increasing and differentiable on an interval I_Λ, possibly unbounded, with $\inf I_\Lambda = 0$. The distribution function Ψ of U is differentiable. The random vector $(C, U, X', Y)'$ satisfies $\Lambda(Y) = \min[\beta' X + U, C]$. Furthermore, U and X are independent, and C and U are conditionally independent given X. The random variable $Z = \beta' X$ has a density p_Z.*

This assumption is needed for equations (3) and (4) to hold. Existence of p_Z is needed for the definitions (5) and (6) to make sense.

Let $S_D \subset R^2$ be a set with the following three properties: if $(z, y) \in S_D$ then $(z, v) \in S_D$ for all $0 \leq v \leq y$; A and B are bounded and continuous on S_D, and the partial derivatives $D_1 A$ and $D_1 B$ exist and are bounded on S_D.

Assumption 2 *The estimation set T_Λ is a Borel subset of I_Λ. By normalization $\Lambda(t_0) = 0$ for some known $t_0 \in T_\Lambda$. The weight function W_Λ is a real function on R^3 such that $W_\Lambda(\cdot, y, t)$ integrates to one for all $t \in T_\Lambda$ and $t_0 \leq y \leq t$ or $t_0 \geq y \geq t$. There is $\epsilon > 0$ such that $B(z, y) > \epsilon$ and $\int_0^y (D_1 A(z, v) B(z, v) - A(z, v) D_1 B(z, v))/B(z, v)^2 \, dv > \epsilon$ for all (z, y) in S_Λ, the union of the supports of $W_\Lambda(\cdot, \cdot, t)$ for $t \in T_\Lambda$, and such that any ball of radius ϵ centered at a point in S_Λ is contained in S_D.*

Assumption 3 *The estimation set T_Ψ is a Borel subset of R. The weight function W_Ψ is a real function on R such that $W_\Psi(\cdot, t)$ integrates to one for all $t \in T_\Psi$ and if $W_\Psi(y, t) \neq 0$ then $y \in T_\Lambda$. There is $\epsilon > 0$ such that $B(\Lambda(y) - t, y) > \epsilon$ for (y, t) in the support S_Ψ of W_Ψ and such that any ball of radius ϵ centered at $(\Lambda(y) - t, y)$ with (y, t) in S_Ψ is a subset of S_D.*

Further assumptions regarding smoothness of the weight functions are listed in assumption 8. Under assumptions 1–3 the right-hand side in equations (8) and (10) are well defined, and the estimators Λ_n and Ψ_n are well defined with probability approaching one as the sample size grows.

The kernel estimators A_n and B_n are based on a sample from the distribution P of the observable random vector (M, X, Y) and a given estimate β_n of β. We need assumptions on the nature of the sample and the estimator β_n.

Assumption 4 *The sequence* $\{(M_i, X_i, Y_i)\}_{i=1}^n$ *is a random sample from* P.

Let P_n denote the empirical measure formed from the n independent observations on P, that is, P_n puts probability $1/n$ on each of the observations. Let β_1 be the first component of β and β_{-1} the vector of remaining components. (Throughout β_1 denotes the first component of β, not the first vector in the sequence β_n.) Similarly, let β_{n1} be the first component of β_n and $\beta_{n,-1}$ the vector of remaining components.

Assumption 5 $|\beta_1| = 1$. *There is a bounded,* P*-measurable function* $\Omega : R^{r+2} \to R^{r-1}$ *such that* $P\Omega = 0$, *the components of* $P\Omega\Omega'$ *are finite and* $n^{1/2}(\beta_{n,-1} - \beta_{-1}) = n^{1/2}P_n\Omega + o_p(1)$ *as* $n \to \infty$.

All the estimators of β mentioned in the introduction satisfy this assumption.

In the following k_z and k_y are integers determined in assumption 6. The next two assumptions concern the choice of bandwidths and kernel functions.

Assumption 6 *The bandwidths* κ_{zn} *and* κ_{yn} *are sequences of positive real numbers converging to 0 such that* $n^{1/2}\kappa_{zn}^{k_z} \to 0$, $n^{1/2}\kappa_{yn}^{k_y} \to 0$, $n^{-1}\kappa_{zn}^{-8} \to 0$ *and* $n^{-1/2}\kappa_{zn}^{-3}\kappa_{yn}^{-1}\log n$ *is bounded.*

Assumption 7 *The kernels* K_z *and* K_y *are bounded, real functions on* R *with supports in* $[-1, 1]$, $\int K_z(s)\,ds = 1$ *and* $\int K_y(s)\,ds = 1$. *In addition,* $\int K_z(s)s^j\,ds = 0$ *for* $j = 1, \ldots, k_z - 1$ *and* $\int K_y(s)s^j\,ds = 0$ *for* $j = 1, \ldots, k_y - 1$. *Furthermore,* K_z *has a derivative*

DK_z which satisfies the Lipschitz condition $|DK_z(z) - DK_z(z^*)| \leq c|z - z^*|$ for some constant c and all $z, z^* \in R$.

The relationship between bandwidth convergence rates and minimum kernel orders were discussed in section 2.2.

The proof of consistency and asymptotic normality is based on Taylor expansions of the integrands in (8) and (10). For the remainder terms and for the bias of the leading terms to vanish sufficiently quickly, we need certain functions to be smooth and bounded. This is taken care of in assumptions 8 and 9.

Assumption 8

1. *The density p_Z is bounded on $I_Z = \{z : (z, y) \in S_D \text{ for some } y\}$.*
2. *The derivatives $D\Lambda$, $D\Lambda^{-1}$ and $D^2\Lambda^{-1}$ exist and are bounded and continuous on $I_Y = \{y : (z, y) \in S_D, \text{some } z\}$.*
3. *The function A is bounded and D_1A, D_1^2A, D_1^3A exist and are bounded on S_D.*
4. *The function B is bounded and D_1B, D_1^2B, D_1^3B, D_2B, D_2D_1B exist and are bounded on S_D.*
5. *The weight function W_Λ is bounded and D_1W_Λ, $D_1^2W_\Lambda$, D_2W_Λ, $D_2D_1W_\Lambda$ and D_3W_Λ exist and are bounded on R^3.*
6. *The weight function W_Ψ is bounded and D_1W_Ψ and D_2W_Ψ exist and are bounded on R^2.*

Let P_1 be the distribution of (M, X_{-1}) and let P_2 be the distribution of (Y, M, X_{-1}), where X_{-1} is the $(r-1)$ last components of X.

Assumption 9 *The conditional distribution of (Z, Y) given $(M, X_{-1}) = (1, x_{-1})$ has a bounded density $\zeta_1(\cdot, \cdot, 1, x_{-1})$ on S_D with respect to Lebesgue measure, and $D_1^i\zeta_1(\cdot, \cdot, 1, x_{-1})$ with $i = 0, 1, \ldots, k_z + 2$ and $D_2^jD_1^i\zeta_1(\cdot, \cdot, 1, x_{-1})$ with $i = 0, 1, 2$ and $j = 0, 1, \ldots, k_y$ exist and are bounded and continuous on S_D almost surely $[P_1(1, x_{-1})]$.*

The conditional distribution of Z given $(Y, M, X_{-1}) = (y, m, x_{-1})$ has a bounded density $\zeta_2(\cdot, y, m, x_{-1})$ on S_D with respect to Lebesgue measure, and $D_1^j D_1^i \zeta_2(\cdot, y, m, x_{-1})$, $i = 0, 1, 2$ and $j = 0, 1, \ldots, k_z$ exist and are bounded and continuous on S_D almost surely $[P_2(y, m, x_{-1})]$.

In addition, X_{-1} is bounded with probability one.

Assumption 9 implies that X_1 is a continuous random variable, but X_{-1} can have discrete components. For β to be identified and the estimators to work there must be a continuous random variable among the components of X, and the corresponding component of β must be known to be nonzero (-1 or 1 after normalization).

Let $\mathcal{X}_\Lambda$ be the set of all bounded, real functions on T_Λ and equip $\mathcal{X}_\Lambda$ with the metric generated by the uniform norm and the σ-algebra generated by closed balls.

Theorem 1 *Under assumptions 1–9 (except 3 and 8.6),*

a. *The sequence $\sup_{t \in T_\Lambda} |\Lambda_n(t) - \Lambda(t)|$ converges to 0 in probability.*

b. *The sequence $\{n^{1/2}(\Lambda_n - \Lambda)\}$ of random elements of $\mathcal{X}_\Lambda$ converges in distribution to a Gaussian stochastic process E_Λ on T_Λ. The mean of E_Λ is zero and the covariance function is $P\lambda(t)\lambda(t^*)$ for $t, t^* \in T_\Lambda$, where $\lambda(t)(\cdot,\cdot,\cdot)$ is a P-measurable, real function on R^3.*

The theorem is proved in appendix A and the function λ is given in equation (38). We show that

$$\sup_{t \in T_\Lambda} |\Lambda_n(t) - \Lambda(t) - P_n\lambda(t)| = o_p(n^{-1/2}) \tag{20}$$

and prove that $P_n\lambda$ as a stochastic process on T_Λ converges in distribution to E_Λ using empirical process methods described by Pollard (1984) and Pakes and Pollard (1989). A method for estimating the covariance function is outlined in section A.3.

The next theorem establishes the limiting behavior of Ψ_n. Let $\mathcal{X}_\Psi$ be the set of all bounded, real functions on T_Ψ and equip $\mathcal{X}_\Psi$ with the metric generated by the uniform norm and the σ-algebra generated by closed balls.

Theorem 2 *Under assumptions 1–9,*

a. *The sequence* $\sup_{t\in T_\Psi}|\Psi_n(t)-\Psi(t)|$ *converges to 0 in probability.*

b. *The sequence* $\{n^{1/2}(\Psi_n-\Psi)\}$ *of random elements of* $\mathcal{X}_\Psi$ *converges in distribution to a Gaussian stochastic process* E_Ψ *on* T_Ψ*. The mean of* E_Ψ *is zero and the covariance function is* $P\psi(t)\psi(t^*)$ *for* $t,t^*\in T_\Psi$*, where* $\psi(t)(\cdot,\cdot,\cdot)$ *is a* P*-measurable, real function on* R^3*.*

The proof of the theorem is found in appendix A and ψ is defined in equation (47). A method for estimating the covariance function is outlined in section A.3.

4 Monte Carlo Results

To illustrate the performance of the estimators in small samples, we now present the results of a few Monte Carlo experiments. In all experiments the data are generated from (2) with $\Lambda(y)=2\log(y/4)$, X and U standard normal scalars and C normally distributed with mean μ_C and unit variance. Since X is a scalar, $\beta=1$ is determined by scale normalization and is not estimated. This enables the experiments to focus on the estimators of Λ and Ψ, which are the objects of interest here. We consider three different degrees of censoring: $\mu_C=\infty$, $\mu_C=1.17$ and $\mu_C=0$ corresponding to expected censoring rates of 0, 20 and 50 per cent. The predicted quantile is the median. The sample size in each experiment is 200 and there are 20 replications per experiment. Other differences among the three experiments reported here are listed in table 1. We have also tried generating data from a non-concave model where Λ is based on the hyperbolic sine function. The results (not reported here) are similar to the model based on the logarithmic function.

In the estimation, the kernel functions used are

$$K_z(x)=\frac{315}{2048}(15-140x^2+378x^4-396x^6+143x^8)1(|x|\le 1) \tag{21}$$

Table 1: Monte Carlo Experiments

Experiment	Expected Censoring	T_Λ	T_Ψ	T_τ	κ_{zn}	κ_{yn}
1	0%	[.6,14]	[-3,2]	[-3,2]	3	1.5
2	25%	[.6,8]	[-3,1.2]	[-3,1]	3	1.5
3	50%	[.8,6]	[-2,1]	[-2.5,.5]	3	2.0

and

$$K_y(x) = \frac{105}{64}(1 - 5x^2 + 7x^4 - 3x^6)1(|x| \leq 1). \tag{22}$$

These are sixth and fourth order kernels, respectively, taken from Müller (1984).

In each experiment the support of $W_\Lambda(\cdot, y, t)$ is the interval where $A(\cdot, y) > \epsilon$, where ϵ is chosen such that on average 95% of the uncensored observations are in the set $\{A > \epsilon\}$. W_Λ itself is constructed as in (15) with

$$f(x) = \frac{15}{16}(1 - x^2)^2 1(|x| \leq 1). \tag{23}$$

Similarly, the support of W_Ψ is chosen such that $W_\Psi(y, t) > 0$ implies $A(\Lambda(y) - t, y) > \epsilon$, and W_Ψ is also formed as in (15) with f as above. More censoring results in a lower probability of large observations, so given the choice of support of weight functions, more censoring shortens the interval on which Λ and Ψ can be estimated. This is reflected in table 1 in the shrinking of T_Λ, T_Ψ and T_τ as the expected censoring increases.

Compared to the formulas (13) and (14) the estimators are modified in the following way. In the event that a denominator is zero we set the entire ratio equal to zero; in finite samples this happens with positive probability. In addition, whenever the inner integral in (14) is smaller than -1 we set it equal to -1.

The results of the experiments are summarized graphically in figures 1–3. The left-hand panels show the average of 20 estimates of Λ, Ψ and $y_{.5}$ (solid lines) and the true value of these functions (dotted lines). The right-hand panels show the first five estimates (solid

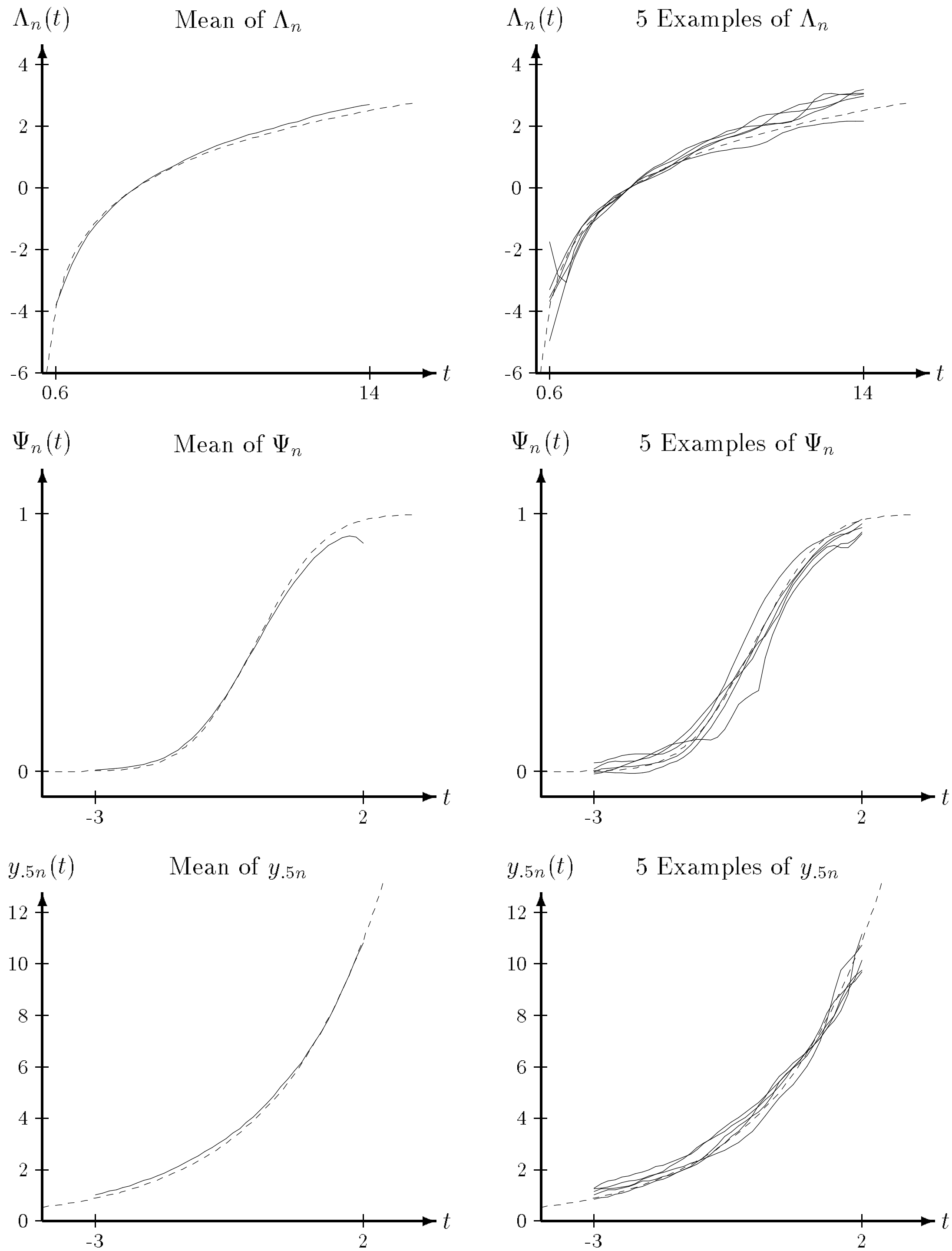

Figure 1: Experiment 1 (No Censoring)

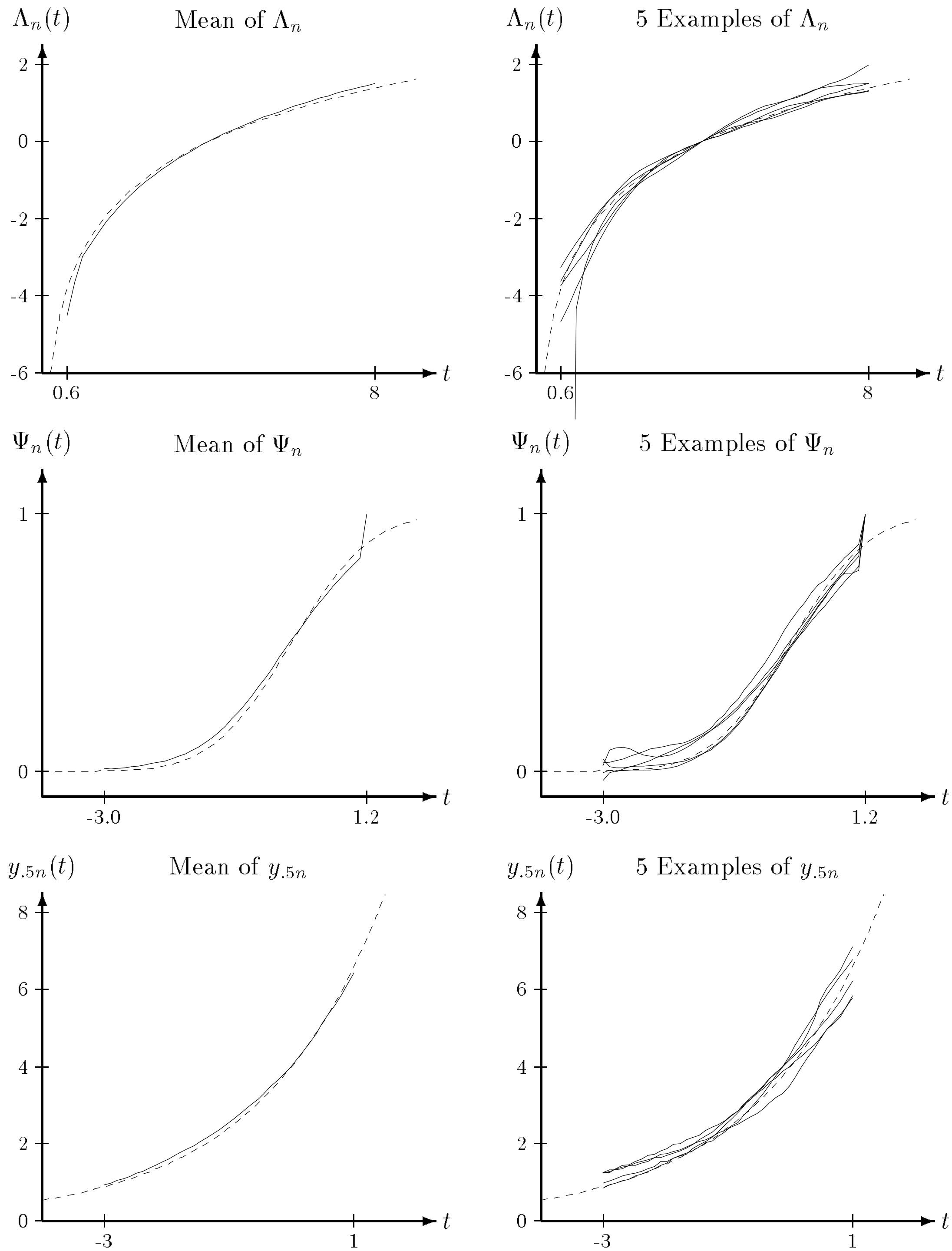

Figure 2: Experiment 2 (25% Censoring)

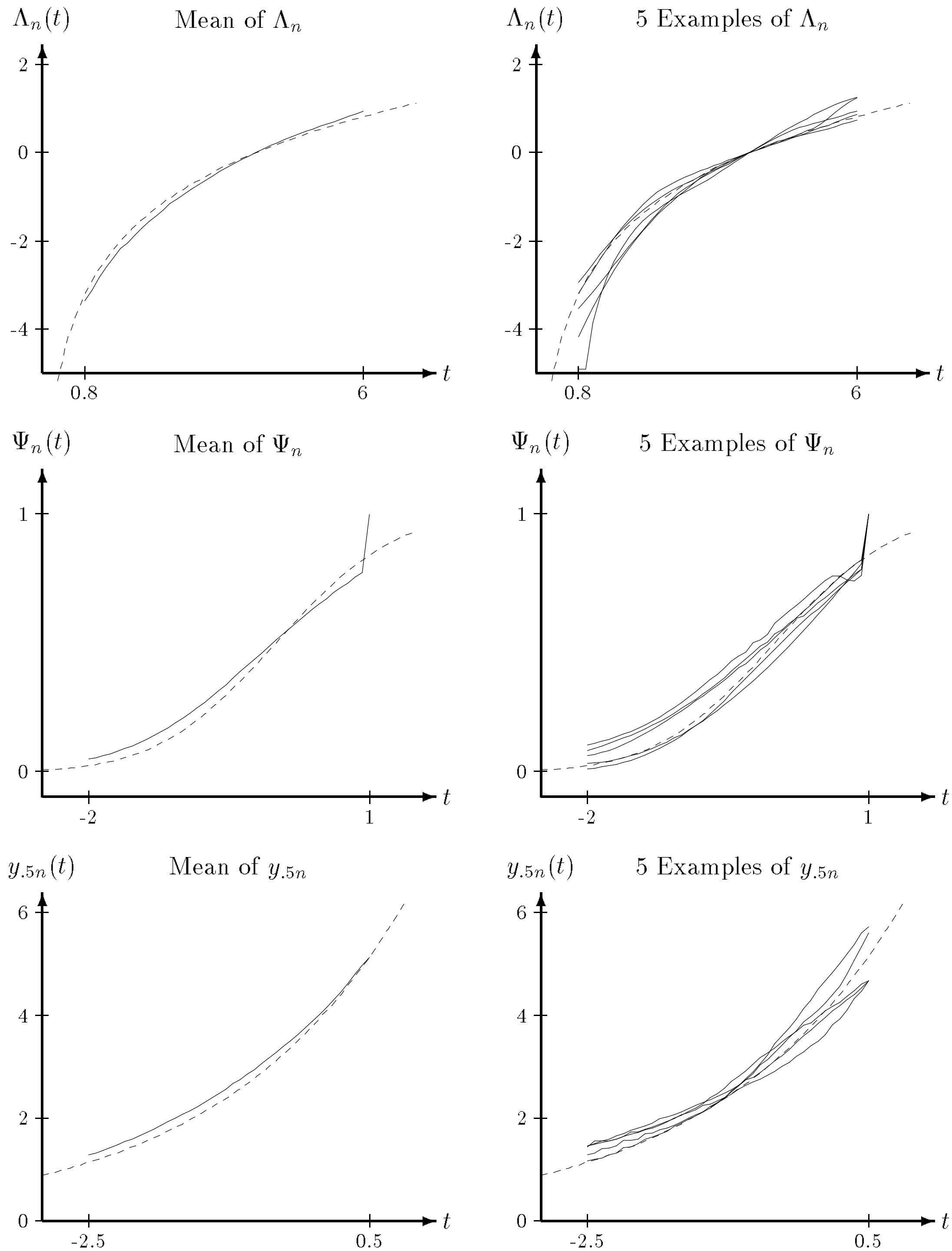

Figure 3: Experiment 3 (50% Censoring)

lines) together with the true values (dotted lines). It can be seen that the new estimators perform well in this setup; even with 50 percent censoring the average of the estimates are very close to the truth. The individual estimates are of course more variable, and in experiment 2 and 3 the jump in all estimates of Ψ at the upper end of T_Ψ is evidence that we may have been too ambitious in selecting T_Ψ. For each experiment we have chosen what we believe are reasonable estimation intervals; of course the estimators will appear less favorable the longer the estimation intervals.

5 Concluding Remarks

In this paper we have proposed new estimators for the censored regression model with an unknown transformation of the dependent variable. We have obtained their asymptotic distributions and presented Monte Carlo evidence that the estimators can perform well in samples of moderate size. However, implementing the estimators is complicated because of the choices that must be made about estimation sets, weight functions, kernel functions and bandwidths. There is plenty of room for further research on practical aspects of implementation.

A Proofs

To simplify the notation, for (z, y) such that $B(z, v) > 0$ for all $v \in [0, y]$ define

$$H(z, y) = \int_0^y \frac{A(z, v)}{B(z, v)} \, dv = -\log[1 - \Psi(\Lambda(y) - z)]. \tag{24}$$

The left-hand side of equation (7) is $D_2 H(z, y)/D_1 H(z, y)$ and the right-hand side of equation (9) is $1 - \exp(-H(z, y))$, the continuous-time equivalent of the Kaplan-Meier estimator of the distribution function of U. On its domain H equals the integrated conditional hazard function of Y^*, because the conditional distribution function G of Y^* at y given $Z = z$ is $G(z, y) = \Psi(\Lambda(y) - z)$.[5] Define also the estimator of H

$$H_n(z, y) = \int_0^y \frac{A_n(z, v)}{B_n(z, v)} \, dv. \tag{25}$$

(If $B_n(z, v) = 0$ for any $0 \le v \le y$ put $H_n(z, y) = 0$.)

A.1 Approximation by Empirical Processes

Let Γ be a real function defined on a set T and let Γ_n be an estimator of Γ defined on T.[6] Let $\mathcal{X}$ be the set of all bounded, real functions on T and equip $\mathcal{X}$ with the metric generated by the uniform norm and the σ-algebra generated by closed balls. Let γ be a function on $R^{r+2} \times T$, and let E be a stochastic process on T such that the sample paths of E are bounded and uniformly continuous and the finite-dimensional distributions of E are jointly normal with zero means and covariance $P\gamma(s)\gamma(t)$ for $s, t \in T$.[7]

[5] The conditional hazard function of Y^* itself is defined on a larger set than H; in particular, the domain of H depends on the censoring scheme which is irrelevant for Y^*.

[6] To ensure measurability it suffices that T is a Borel subset of a compact metric space, see Pakes and Pollard (1989, p1031) or appendix C of Pollard (1984).

[7] Notation: P is a probability measure on R^{r+2}, and Γ_n and γ are real functions $R^{r+2} \times T$. The dependence on the probability space is often suppressed, and $\gamma(t)$ is short for the random variable $\gamma(\cdot, t)$. Similarly, $P\gamma(t)$ denotes the expected value of $\gamma(t)$, while $P\gamma$ is a function on T whose value at t is $(P\gamma)(t) = P\gamma(t)$, and $P_n\gamma$ is the stochastic process on T defined by $(P_n\gamma)(t) = P_n\gamma(t) = (1/n)\sum_{i=1}^n \gamma(M_i, X_i, Y_i, t)$.

Lemma 1 *If the class $\{\gamma(t) : t \in T\}$ is permissible and Euclidean for a constant envelope,*[8] *$P\gamma(t) = 0$ for all $t \in T$ and $\sup_{t\in T}|\Gamma_n(t) - \Gamma(t) - P_n\gamma(t)| = o_p(n^{-1/2})$, then $\sup_{t\in T}|\Gamma_n(t) - \Gamma(t)| \to 0$ in probability and $n^{1/2}(\Gamma_n - \Gamma) \to E$ in distribution as random elements of $\mathcal{X}$.*

Proof The first conclusion of the lemma follows from theorem II.24 of Pollard (1984) (same as lemma 2.8 of Pakes and Pollard 1989), and the second from lemma 2.16 of Pakes and Pollard (1989) and theorem VII.21 of Pollard (1984). ■

Theorems 1 and 2 follow from lemma 1 if we can find functions λ and ψ which satisfy the assumptions of the lemma with $\Gamma_n - \Gamma = \Lambda_n - \Lambda$ and $\Gamma_n - \Gamma = \Psi_n - \Psi$, respectively. Both Λ_n and Ψ_n are integrals, and the first step is to linearize the integrands in the kernel functions.

Consider first Λ_n. With the conventions from footnote 3, define on $R^3 \times T_\Lambda$

$$f_1(z,y,m,t) = -\frac{W_\Lambda(z,y,t)[1(t_0 \le y \le t) - 1(t \le y \le t_0)]m}{B(z,y)D_1H(z,y)}, \tag{26}$$

$$f_2(z,y,m,t) = -\int_{t_0}^{t} \frac{W_\Lambda(z,v,t)1(y > v)D\Lambda(v)}{B(z,v)}\,dv, \tag{27}$$

$$f_3(z,y,m,t) = \frac{D_1B(z,y)m}{B(z,y)^2}\int_{t_0}^{t} \frac{W_\Lambda(z,v,t)1(v \ge y \ge 0)D\Lambda(v)}{D_1H(z,v)}\,dv, \tag{28}$$

$$f_4(z,y,m,t) = -\frac{m}{B(z,y)}\int_{t_0}^{t} \frac{W_\Lambda(z,v,t)1(v \ge y \ge 0)D\Lambda(v)}{D_1H(z,v)}\,dv, \tag{29}$$

$$f_5(z,y,m,t) = \int_{t_0}^{t}\Bigg[\frac{W_\Lambda(z,v,t)D\Lambda(v)}{D_1H(z,v)} \times \int_0^v 1(y > \xi)\frac{D_1A(z,\xi)B(z,\xi) - 2A(z,\xi)D_1B(z,\xi)}{B(z,\xi)^3}\,d\xi\Bigg]dv, \tag{30}$$

$$f_6(z,y,m,t) = \int_{t_0}^{t}\Bigg[\frac{W_\Lambda(z,v,t)D\Lambda(v)}{D_1H(z,v)}\int_0^v 1(y > \xi)\frac{A(z,\xi)}{B(z,\xi)^2}\,d\xi\Bigg]dv. \tag{31}$$

[8] See Pakes and Pollard (1989).

By definition and Fubini's theorem, for $t \in T_\Lambda$

$$
\begin{aligned}
\Lambda_n(t) - \Lambda(t) &= \frac{1}{n}\sum_{i=1}^n \iint f_1(z,y,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\frac{1}{\kappa_{yn}}K_y\left(\frac{Y_i - y}{\kappa_{yn}}\right)dy\,dz \\
&+ \frac{1}{n}\sum_{i=1}^n \int f_2(z,Y_i,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)dz \\
&+ \frac{1}{n}\sum_{i=1}^n \iint f_3(z,y,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\frac{1}{\kappa_{yn}}K_y\left(\frac{Y_i - y}{\kappa_{yn}}\right)dy\,dz \\
&+ \frac{1}{n}\sum_{i=1}^n \iint f_4(z,y,M_i,t)\frac{1}{\kappa_{zn}^2}DK_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)\frac{1}{\kappa_{yn}}K_y\left(\frac{Y_i - y}{\kappa_{yn}}\right)dy\,dz \\
&+ \frac{1}{n}\sum_{i=1}^n \int f_5(z,Y_i,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)dz \\
&+ \frac{1}{n}\sum_{i=1}^n \int f_6(z,Y_i,M_i,t)\frac{1}{\kappa_{zn}^2}DK_z\left(\frac{\beta_n' X_i - z}{\kappa_{zn}}\right)dz \\
&+ R_n^\Lambda(t),
\end{aligned}
\tag{32}
$$

where the remainder term is the formidable expression

$$
\begin{aligned}
R_n^\Lambda(t) &= \iint_{t_0}^t W_\Lambda(z,y,t)\frac{A(z,y)D_1H_n(z,y)\big(B_n(z,y)-B(z,y)\big)^2}{B_n(z,y)B(z,y)^2 D_1H(z,y)^2}\,dy\,dz \\
&+ \iint_{t_0}^t W_\Lambda(z,y,t)\frac{2A(z,y)\big(B_n(z,y)-B(z,y)\big)\big(D_1H_n(z,y)-D_1H(z,y)\big)}{B_n(z,y)B(z,y)D_1H(z,y)^2}\,dy\,dz \\
&+ \iint_{t_0}^t W_\Lambda(z,y,t)\frac{A(z,y)\big(D_1H_n(z,y)-D_1H(z,y)\big)^2}{B_n(z,y)D_1H_n(z,y)D_1H(z,y)^2}\,dy\,dz \\
&- \iint_{t_0}^t W_\Lambda(z,y,t)\frac{\big(A_n(z,y)-A(z,y)\big)\big(B_n(z,y)-B(z,y)\big)}{B_n(z,y)B(z,y)D_1H(z,y)}\,dy\,dz \\
&- \iint_{t_0}^t W_\Lambda(z,y,t)\frac{\big(A_n(z,y)-A(z,y)\big)\big(D_1H_n(z,y)-D_1H(z,y)\big)}{B_n(z,y)D_1H_n(z,y)D_1H(z,y)}\,dy\,dz \\
&- \iint_{t_0}^t W_\Lambda(z,y,t)\frac{A(z,y)\big(B_n(z,y)-B(z,y)\big)\big(D_1H_n(z,y)-D_1H(z,y)\big)}{B(z,y)^2 D_1H(z,y)^2}\,dy\,dz \\
&+ \iint_{t_0}^t \int_0^y W_\Lambda(z,y,t)\frac{A(z,y)}{B(z,y)^2 D_1H(z,y)^2}D_1B(z,\xi)\big(B_n(z,\xi)+B(z,\xi)\big) \\
&\quad \times \frac{\big(A_n(z,\xi)-A(z,\xi)\big)\big(B_n(z,\xi)-B(z,\xi)\big)}{B(z,\xi)^2 B_n(z,\xi)^2}\,d\xi\,dy\,dz \\
&+ \iint_{t_0}^t \int_0^y W_\Lambda(z,y,t)\frac{A(z,y)}{B(z,y)^2 D_1H(z,y)^2}
\end{aligned}
$$

$$
\begin{aligned}
&\times \frac{D_1 A(z,\xi) B(z,\xi) B_n(z,\xi) - A(z,\xi) D_1 B(z,\xi)(2B_n(z,\xi) + B(z,\xi))}{B(z,\xi)^3 B_n(z,\xi)^2} \\
&\times (B_n(z,\xi) - B(z,\xi))^2 \, d\xi \, dy \, dz \\
- &\iint_{t_0}^{t} \int_0^y W_\Lambda(z,y,t) \frac{A(z,y)}{B(z,y)^2 D_1 H(z,y)^2} \\
&\times \frac{(A_n(z,\xi) - A(z,\xi))(D_1 B_n(z,\xi) - D_1 B(z,\xi))}{B_n(z,\xi)^2} \, d\xi \, dy \, dz \\
- &\iint_{t_0}^{t} \int_0^y W_\Lambda(z,y,t) \frac{A(z,y)}{B(z,y)^2 D_1 H(z,y)^2} \\
&\times \frac{B_n(z,\xi)(B_n(z,\xi) - B(z,\xi))(D_1 A_n(z,\xi) - D_1 A(z,\xi))}{B(z,\xi) B_n(z,\xi)^2} \, d\xi \, dy \, dz \\
+ &\iint_{t_0}^{t} \int_0^y W_\Lambda(z,y,t) \frac{A(z,y)}{B(z,y)^2 D_1 H(z,y)^2} \frac{A(z,\xi)(B_n(z,\xi) + B(z,\xi))}{B(z,\xi)^2 B_n(z,\xi)^2} \\
&\times (B_n(z,\xi) - B(z,\xi))(D_1 B_n(z,\xi) - D_1 B(z,\xi)) \, d\xi \, dy \, dz. \qquad (33)
\end{aligned}
$$

The integrands on the right-hand side of (33) are zero outside $(z,y) \in S_\Lambda$ and $(z,\xi) \in S_\Lambda^*$, where $S_\Lambda^* \subset R^2$ is the set of (z,ξ) such that $0 \le \xi$ and $(z,y) \in S_\Lambda$ for some $\xi \le y$. We argue in section A.2 that

$$
\sup_{(z,y) \in S_\Lambda^*} \left| D_1^d A_n(z,y) - D_1^d A(z,y) \right| = o_p(n^{-1/4}) \qquad (34)
$$

and

$$
\sup_{(z,y) \in S_\Lambda^*} \left| D_1^d B_n(z,y) - D_1^d B(z,y) \right| = o_p(n^{-1/4}), \qquad (35)
$$

where $d = 0$ or $d = 1$. By assumption 8 $D_1^d A$ and $D_1^d B$ are bounded on S_D and by construction there is $\epsilon > 0$ such that $B > \epsilon$ on S_Λ^*. Uniform boundedness from below of B and uniform convergence of B_n imply that $B_n > \epsilon/2$ on S_Λ^* with probability approaching one for large n. Therefore, $\left| D_i^d H_n - D_i^d H \right| = o_p(n^{-1/4})$ uniformly on S_Λ for $i = 1, 2$ and $d = 0, 1$. By construction $D_1 H > \epsilon$ on S_Λ, and again uniform boundedness from below and uniform convergence imply that $D_1 H_n > \epsilon/2$ on S_Λ with probability approaching one for large n. Therefore, each integrand in equation (33) is a product of a uniformly bounded term and two terms which converge uniformly to zero at the rate $o_p(n^{-1/4})$. It follows that $\sup_{t\in T_\Lambda}|R_n^\Lambda(t)| = o_p(n^{-1/2})$.

For $i = 1, 3$ with $d = 0$ and for $i = 4$ with $d = 1$ define Φ_i on $R^{r+2} \times T_\Lambda$ by

$$\begin{aligned}\Phi_i(m, x, y, t) &= D_1^d f_i(\beta' x, y, m, t) \\ &- \Omega(m, x, y)' \iiint f_i(z, y^*, m^*, t) x_{-1}^* D_1^{d+1} \zeta_1(z, y^*, m^*, x_{-1}^*)\, dz\, dy^*\, dP_1(m^*, x_{-1}^*), \qquad (36)\end{aligned}$$

where Ω is given in assumption 5 and ζ_1 and P_1 are defined in assumption 9. Similarly, for $i = 2, 5$ with $d = 0$ and for $i = 6$ with $d = 1$ define Φ_i on $R^{r+2} \times T_\Lambda$ by

$$\begin{aligned}\Phi_i(m, x, y, t) &= D_1^d f_i(\beta' x, y, m, t) \\ &- \Omega(m, x, y)' \iint f_i(z, y^*, m^*, t) x_{-1}^* D_1^{d+1} \zeta_2(z, y^*, m^*, x_{-1}^*)\, dz\, dP_2(y^*, m^*, x_{-1}^*), \qquad (37)\end{aligned}$$

where ζ_2 and P_2 are defined in assumption 9. Finally, define λ on $R^{r+2} \times T_\Lambda$ by

$$\lambda(m, x, y, t) = \sum_{i=1}^{6} \Phi_i(m, x, y, t). \qquad (38)$$

It can be verified that $P\lambda(t) = 0$ for all $t \in T_\Lambda$. Suppose the leading term in (32) involving f_i can be approximated uniformly to the order $o_p(n^{-1/2})$ by the empirical process $P_n\Phi_i$ and that the classes $\{\Phi_i(t) : t \in T_\Lambda\}$ are permissible and Euclidean for constant envelopes. (This is proved in section A.2.) Then by lemma 2.14(i) of Pakes and Pollard (1989)

$$\sup_{t\in T_\Lambda} |\Lambda_n(t) - \Lambda(t) - P_n\lambda(t)| = o_p(n^{-1/2}), \qquad (39)$$

and by lemma 2.14 of Pakes and Pollard (1989) the class $\{\lambda(t) : t \in T_\Lambda\}$ is permissible and Euclidean. The assumptions of lemma 1 are therefore satisfied with $\Gamma_n - \Gamma = \Lambda_n - \Lambda$, $\gamma = \lambda$ and $T = T_\Lambda$ and theorem 1 follows.

Turning now to Ψ_n, define on $R^3 \times T_\Psi$

$$f_7(v,y,m,t) = W_\Psi(v,t)[1-\Psi(t)]\,\frac{1(0 \le y \le v)m}{B(\Lambda(v)-t,y)} \tag{40}$$

and

$$f_8(v,y,m,t) = -W_\Psi(v,t)[1-\Psi(t)]\int_0^v \frac{1(y>\xi)A(\Lambda(v)-t,\xi)}{B(\Lambda(v)-t,\xi)^2}\,d\xi. \tag{41}$$

By the mean value theorem $\exp(-H_n)-\exp(-H) = -\exp(-H_n^*)(H_n-H)$, where H_n^* is between H and H_n, so for $t \in T_\Psi$

$$\begin{aligned}\Psi_n(t)-\Psi(t) = {} & \frac{1}{n}\sum\iint f_7(v,y,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i-\Lambda_n(v)+t}{\kappa_{zn}}\right)\frac{1}{\kappa_{yn}}K_y\left(\frac{Y_i-y}{\kappa_{yn}}\right)dy\,dv \\ & + \frac{1}{n}\sum\int f_8(v,Y_i,M_i,t)\frac{1}{\kappa_{zn}}K_z\left(\frac{\beta_n' X_i-\Lambda_n(v)+t}{\kappa_{zn}}\right)dv \\ & + R_n^\Psi(t),\end{aligned} \tag{42}$$

where

$$\begin{aligned}R_n^\Psi(t) = {} & \int W_\Psi(v,t)\Big[\exp\big(-H_n^*(\Lambda_n(v)-t,v)\big)-\exp\big(-H(\Lambda(v)-t,v)\big)\Big] \\ & \quad \times \big[H_n(\Lambda_n(v)-t,v)-H(\Lambda(v)-t,v)\big]\,dv \\ & - \iint_0^v W_\Psi(v,t)[1-\Psi(t)] \\ & \quad \times \Bigg[\frac{\big(A_n(\Lambda_n(v)-t,y)-A(\Lambda(v)-t,y)\big)\big(B_n(\Lambda_n(v)-t,y)-B(\Lambda(v)-t,y)\big)}{B(\Lambda(v)-t,y)B_n(\Lambda_n(v)-t,y)} \\ & \quad - \frac{A(\Lambda(v)-t,y)\big(B_n(\Lambda_n(v)-t,y)-B(\Lambda(v)-t,y)\big)^2}{B(\Lambda(v)-t,y)^2 B_n(\Lambda_n(v)-t,y)}\Bigg]dy\,dv.\end{aligned}$$

In section A.2 we argue that

$$\sup_{\substack{(v,t)\in S_\Psi \\ 0\le y\le v}} \big|A_n(\Lambda_n(v)-t,y)-A(\Lambda(v)-t,y)\big| = o_p(n^{-1/4}) \tag{43}$$

and

$$\sup_{\substack{(v,t)\in S_\Psi \\ 0\le y\le v}} \left|B_n(\Lambda_n(v)-t,y)-B(\Lambda(v)-t,y)\right| = o_p(n^{-1/4}). \tag{44}$$

This and the defining properties of S_Ψ imply that $B(\Lambda(v)-t,y)>\epsilon$ and $B_n(\Lambda_n(v)-t,y)>\epsilon/3$ with high probability for large n for $(v,t)\in S_\Psi$, $0\le y\le v$ and some $\epsilon>0$. It follows that $\left|H_n(\Lambda_n(v)-t,y)-H(\Lambda(v)-t,y)\right| = o_p(n^{-1/4})$ uniformly. The integrands in (43) are products of a uniformly bounded term and two terms which convergence uniformly at the rate $o_p(n^{-1/4})$, and therefore $\sup_{t\in T_\Psi}\left|R_n^\Psi(t)\right| = o_p(n^{-1/2})$.

Define Φ_7 on $R^{r+2}\times T_\Psi$ by

$$\begin{aligned}
\Phi_7(m,x,y,t) &= f_7(\Lambda^{-1}(\beta'x+t),y,m,t)D\Lambda^{-1}(\beta'x+t)\\
&\quad - \Omega(m,x,y)'\iiint f_7(v,y^*,m^*,t)x^*_{-1}D_1\zeta_1(\Lambda(v)-t,y^*,m^*,x^*_{-1})\,dv\,dy^*\,dP_1(m^*,x^*_{-1})\\
&\quad + \iiint f_7(v,y^*,m^*,t)\lambda(m,x,y,v)D_1\zeta_1(\Lambda(v)-t,y^*,m^*,x^*_{-1})\,dv\,dy^*\,dP_1(m^*,x^*_{-1}),
\end{aligned} \tag{45}$$

Φ_8 on $R^{r+2}\times T_\Psi$ by

$$\begin{aligned}
\Phi_8(m,x,y,t) &= f_8(\Lambda^{-1}(\beta'x+t),y,m,t)D\Lambda^{-1}(\beta'x+t)\\
&\quad - \Omega(m,x,y)'\iint f_8(v,y^*,m^*,t)x^*_{-1}D_1\zeta_2(\Lambda(v)-t,y^*,m^*,x^*_{-1})\,dv\,dP_2(y^*,m^*,x^*_{-1})\\
&\quad + \iint f_8(v,y^*,m^*,t)\lambda(m,x,y,v)D_1\zeta_2(\Lambda(v)-t,y^*,m^*,x^*_{-1})\,dv\,dP_2(y^*,m^*,x^*_{-1})
\end{aligned} \tag{46}$$

and ψ on $R^{r+2}\times T_\Psi$ by

$$\psi(m,x,y,t) = \Phi_7(m,x,y,t)+\Phi_8(m,x,y,t). \tag{47}$$

Suppose the class $\{\psi(t) : t \in T\}$ is permissible and Euclidean and

$$\sup_{t \in T_\Psi} \left| \Psi_n(t) - \Psi(t) - P_n \psi(t) \right| = o_p(n^{-1/2}). \tag{48}$$

Since $P\psi(t) = 0$ for all $t \in T_\Psi$, the assumptions of lemma 1 are therefore satisfied with $\Gamma_n - \Gamma = \Psi_n - \Psi$, $\gamma = \psi$ and $T = T_\Psi$, and theorem 2 follows.

A.2 Technical Lemmas

In this section we deal with the suppositions made in the previous section. There were a number of those: equations (34) and (35) with $d = 0$ and $d = 1$, equations (43) and (44), and the claims that the leading terms in (32) and (42) involving f_i can be approximated by the empirical processes $P_n \Phi_i$ and that the classes $\{\Phi_i(t) : t \in T\}$ are permissible and Euclidean where $T = T_\Lambda$ or $T = T_\Psi$. We prove only half of the claims (those which do not involve K_y); the omitted proofs are similar.[9] The main results of this section are lemma 7, which implies equations (35) and (44), and lemma 10, which implies that leading terms in (32) and (42) involving f_2, f_5, f_6 and f_8 can be uniformly approximated by the empirical processes $P_n \Phi_i$. Lemma 10 also implies that the classes $\{\Phi_i(t) : t \in T\}$ are permissible and Euclidean.

A little extra notation is required. Let T be a Borel set, let f be a real function on $R^{r+2} \times T$, and let $S_f \subset R$ be the support of $f(\cdot, y, m, x_{-1}, t)$. Let π be a real map from $S_f \times T$ into I_Z, and let π_n be a sequence of random estimators of π. For $(y, m, x_{-1}) \in R^{r+1}$, $x \in R^r$, $s \in S_f$ and $t \in T$ define

$$\phi_n^d(s, t, m, x, y) = f(s, y, m, x_{-1}, t) \frac{D^d K_z(\kappa_{zn}^{-1}(\beta_n' x - \pi_n(s, t)))}{\kappa_{zn}^{1+d}} \tag{49}$$

[9] Interested readers can obtain generalized versions of the lemmas presented in this section from the authors. These lemmas apply to all claims. In addition, their proofs are more detailed.

and for $s \in S_f$ and $t \in T$ define

$$\phi^d(s,t) = \int f(s,y,m,x_{-1},t) D_1^d \zeta_2(\pi(s,t),y,m,x_{-1})\, dP_2(y,m,x_{-1}). \tag{50}$$

Note that the superscript d is part of the name of the functions ϕ_n^d and ϕ^d and not an exponent. Recall that

$$D_1^d B_n(z,v) = \frac{1}{n}\sum_{i=1}^{n} 1(Y_i > v)\frac{1}{\kappa_{zn}} D_1^d K_z\Big(\frac{\beta_n' X_i - z}{\kappa_{zn}}\Big), \tag{51}$$

and a little manipulation of the definition (6) of B reveals that

$$D_1^d B(z,v) = \int 1(y > v) D_1^d \zeta_2(z,y,m,x_{-1})\, dP_2(y,m,x_{-1}). \tag{52}$$

With $S_f = \{z : (z,y) \in S_\Lambda^* \text{ some } y\}$, $s = z$, $T = \{y : (z,y) \in S_\Lambda^* \text{ some } z\}$, $t = v$, $f(s,y,m,x_{-1},t) = 1(y > v)$, $\pi(s,t) = z$ and $\pi_n(s,t) = z$, then $P_n\phi_n^d(s,t) = D_1^d B_n(z,v)$ and $P_n\phi^d(s,t) = D_1^d B(z,v)$. If instead $T = \{y : (z,y) \in S_\Lambda^* \text{ some } z\} \times T_\Psi$, $t = (v,t^*)$, $\pi(s,t) = \Lambda(z) - t^*$ and $\pi(s,t) = \Lambda_n(z) - t^*$, then $P_n\phi_n^d(s,t) = D_1^d B_n(\Lambda_n(z) - t^*, v)$ and $P_n\phi^d(s,t) = D_1^d B(\Lambda(z) - t^*, v)$, which appear in the formulae for Ψ_n. Therefore, if

$$\sup |P_n\phi_n^d - \phi^d| = o_p(n^{-1/4}) \tag{53}$$

then equations (35) and (44) are proved. Equation (53) is proved in lemma 7.

Now let Π be a map from $S_f \times T \times R^{r+2}$ into I_Z such that $\sup|\pi_n - \pi - P_n\Pi| = o_p(n^{-1/2})$, and suppose that $\pi(\cdot,t)$ is invertible, and let $\pi^{-1}(\cdot,t)$ denote the inverse. Define a on $R^{r+2} \times T$ by

$$a(s,y,m,x_{-1},t) = 1_{\pi(S_f,t)}(s) f(\pi^{-1}(s,t),y,m,x_{-1},t) D_1\pi^{-1}(s,t). \tag{54}$$

Then define Φ_n^d on $R^{r+2} \times T$ by

$$\Phi_n^d(m, x, y, t) = \int_{S_f} \phi_n^d(s, t, m, x, y)\, ds \tag{55}$$

and Φ^d on $R^{r+2} \times T$ by

$$\begin{aligned} &\Phi^d(m, x, y, t) = D_1^d a(\beta' x, y, m, x_{-1}, t) \\ &- \Omega(m, x, y)' \iint f(s, y^*, m^*, x_{-1}^*, t) x_{-1}^* D_1^{d+1} \zeta_2(\pi(s, t), y^*, m^*, x_{-1}^*)\, ds\, dP_2(y^*, m^*, x_{-1}^*) \\ &+ \iint f(s, y^*, m^*, x_{-1}^*, t) \Pi(s, t, m, x, y) D_1^{d+1} \zeta_2(\pi(s, t), y^*, m^*, x_{-1}^*)\, ds\, dP_2(y^*, m^*, x_{-1}^*). \end{aligned} \tag{56}$$

Again the superscript d is part of the name of the functions and not an exponent. Let f be either f_2, f_5, f_6 or f_8, and put $T = T_\Lambda$, $\pi(s, t) = s$, $\pi_n(s, t) = s$ and $\Pi(s, t, m, x, y) = 0$ if f is f_2, f_5 or f_6 and put $T = T_\Psi$, $\pi(s, t) = \Lambda(s) - t$, $\pi_n(s, t) = \Lambda(s) - t$ and $\Pi(s, t, m, x, y) = \lambda(m, x, y, s)$ if f is f_8. Then $P_n \Phi_n^d$ equals the leading term in (32) or (42) involving f_i, and Φ^d is the corresponding function Φ_i^d. We show in lemma 10 that $\{\Phi^d(t) : t \in T\}$ is a permissible and Euclidean class for a constant envelope and that

$$\sup \bigl| P_n(\Phi_n^d - \Phi^d) \bigr| = o_p(n^{-1/2}), \tag{57}$$

and (half of) equations (39) and (48) follow.

The path to lemmas 7 and 10 begins by establishing certain properties of f and π. The following lemmas 3–10 are formulated for general f and π satisfying the conclusions of lemma 2.

Lemma 2 *Let $S_a(y, m, x_{-1}, t)$ be the interior of the support of $a(\cdot, y, m, x_{-1}, t)$. For the specifications of T, f, S_f, d and π given in the text we have:*

1. *$f(s, \cdot, \cdot, \cdot, t)$ is uniformly bounded almost surely $[P_2]$ for all $s \in S_f$ and $t \in T$.*
2. *The class $\{f(s, \cdot, \cdot, \cdot, t) : s \in S_f, t \in T\}$ is Euclidean.*

3. *S_f is bounded.*

4. *There is a real function Π on $S_f \times T \times R^{r+2}$ into I_Z such that $\sup|\pi_n - \pi - P_n\Pi| = o_p(n^{-1/2})$ and $\sup|P_n\Pi| = O_p(n^{-1/2})$.*

5. *$D_1 a(\cdot, y, m, x_{-1}, t), \ldots, D_1^d a(\cdot, y, m, x_{-1}, t)$ exist and are bounded on $S_a(y, m, x_{-1}, t)$. There is a number c_0 such that given any $\epsilon > 0$ the boundary of $S_a(y, m, x_{-1}, t)$ can be covered by the union of c_0/ϵ^{q-1} balls of radius ϵ centered on the boundary for all $t \in T$ and almost surely $[P_2(y, m, x_{-1})]$, and there are constants $c_1 \geq 0$ and $\delta_1 \in (0, 1)$ such that for all $t \in T$*

$$\left|D_1^d a(s, y, m, x_{-1}, t) - D_1^d a(s^*, y, m, x_{-1}, t)\right| \leq c_1 |s - s^*|^{\delta_1} \quad \text{a.s. } [P_2(y, m, x_{-1})]$$

whenever $s, s^* \in \pi(S_a(y, m, x_{-1}, t), t)$ and $|s - s^*| \leq 1$.

6. *The class $\{D_1^d a(\cdot, \cdot, \cdot, \cdot, t) : t \in T\}$ is Euclidean, and there are constants $c_2 \geq 0$ and $\delta_2 > 0$ such that*

$$\int_{[-1,1]} \left|D_1^d a(s - \epsilon v, y, m, x_{-1}, t) - D_1^d a(s - \epsilon v, y, m, x_{-1}, t^*)\right| dv \leq c_2 |t - t^*|^{\delta_2}$$

for some $\epsilon > 0$ whenever $t, t^* \in T$.

7. *$D_2\Pi$, $D_1^{d+1}\zeta_2$ and $D_2\pi$ exist and are bounded, and there are constants $c_3 \geq 0$ and $\delta_3 > 0$ such that*

$$\int \left|f(s, y, m, x_{-1}, t) - f(s, y, m, x_{-1}, t^*)\right| dv \leq c_3 |t - t^*|^{\delta_3}$$

whenever $t, t^* \in T$.

Lemmas 9 and 10 are proved for the case where $a(\cdot, y, m, x_{-1}, t)$ may be discontinuous, and part of the point of lemma 2.5 is that all discontinuity points of $a(\cdot, y, m, x_{-1}, t)$ are located on the boundary of $S_a(y, m, x_{-1}, t)$. This may seem an overkill since neither f_2, f_5, f_6 nor f_8 are discontinuous in their first argument. However, f_1 and f_7 are not continuous in the relevant arguments, and we formulate lemmas 9 and 10 for the general case so that they may be easier modified to cover the leading terms in (32) and (42) involving f_1 and f_7.

Proof The classes $\{f_i(z,\cdot,\cdot,t) : z \in I_Z, t \in T_\Lambda\}$, $i = 2, 5, 6, 8$, are Euclidean by lemma 2.13 of Pakes and Pollard (1989), because these functions are Lipschitz continuous in z and t under assumption 8. S_f is bounded by assumption 2 or 3 since the requirement $B > \epsilon$ implies that S_Λ and S_Ψ are bounded. Part 4 follows by theorem 1, since either $\pi(s,t) = s$, $\pi_n(s,t) = s$ and $\Pi(s,t,m,x,y) = 0$ or $\pi(s,t) = \Lambda(s) - t$, $\pi_n(s,t) = \Lambda(s) - t$ and $\Pi(s,t,m,x,y) = \lambda(m,x,y,s)$. Let a_i be the a-function corresponding to f_i. The classes $\{D_1^d a_i(t) : t \in T_\Lambda\}$, $i = 2, 5, 6$, are Euclidean by lemma 2.13 of Pakes and Pollard (1989) because f_2, f_5 and $D_1 f_6$ are integrals from t_0 to t of bounded integrands, and $\{a_8(t) : t \in T_\Psi\}$ is Euclidean by lemmas 2.13 and 2.14(ii) of Pakes and Pollard (1989) because a_8 is the product of an indicator function of a rectangle and a function which has a bounded partial derivative with respect to t. The Lipschitz properties assumed of f and $D_1^d a$ follow from assumption 8 and the differentiability of π, ζ_2 and Π follow from assumptions 8 and 9 and theorem 1. ■

Next, consider the case where β and π are known and define

$$\varphi_n^d(s,t,m,x,y) = f(s,y,m,x_{-1},t)\frac{D^d K_z(\kappa_{zn}^{-1}(\beta' x - \pi(s,t)))}{\kappa_{zn}^{1+d}}, \tag{58}$$

that is, φ_n^d differs from ϕ_n^d in that β and π replace β_n and π_n.

Lemma 3 *If $q = d$ or $q = d+1$ then* $\sup\left|P\varphi_n^q - \phi^q\right| = O(\kappa_{zn}^{k_z})$.

Proof By change of variables and repeated integration by parts

$$P\varphi_n^q(s,t) = \iint f(s,y,m,x_{-1},t) K_z(u) D_1^q \zeta_2(\pi(s,t) + \kappa_{zn} u, y, m, x_{-1})\, du\, dP_2(y,m,x_{-1}).$$

By a Taylor expansion of $D_1^q\zeta_2(\pi(s,t) + \kappa_{zn}u, y, m, x_{-1})$ about $u = 0$ and using the assumption that $\int K_z(u)\,du = 1$ and $\int K_z(u)u^j\,du = 0$ for $j = 1, \ldots, k_z - 1$, it follows that

$$P\varphi_n^q(s,t) = \phi^q(s,t) + \kappa_{zn}^{k_z} \iint f(s,y,m,x_{-1},t) K_z(u) \frac{D_1^{q+k_z}\zeta_2(\pi(s,t) + \kappa_{zn}u^*, y, m, x_{-1})}{k_z!} \times u^{k_z}\,du\,dP_2(y,m,x_{-1}),$$

where u^* is between 0 and u. The lemma follows. ∎

Lemma 4 *Let ρ be a real function of bounded variation on R. If $S_f \times T$ is a Borel subset of a Euclidean space and $\{f(s,\cdot,\cdot,\cdot,t) : s \in S_f, t \in T\}$ is a permissible, Euclidean class for a constant envelope, then the class of functions of the form $(m,x,y) \mapsto f(s,y,m,x_{-1},t)\rho(\kappa_{zn}^{-1}(\beta' x - \pi(s,t)))$ for $s \in S_f$ and $t \in T$ is permissible and Euclidean for a constant envelope.*

Proof By lemma 22 of Nolan and Pollard (1987) the class of all functions on R of the form $(m,x,y) \mapsto \rho(\kappa_{zn}^{-1}(\beta' x - u))$ with $u \in R$ is Euclidean for a constant envelope, and it is a permissible class because $S_f \times T$ is a Borel subset of a Euclidean space, see appendix C of Pollard (1984). By the definition of a Euclidean class replacing u by $\pi(s,t)$ and indexing by $(s,t) \in S_f \times T$ instead of $u \in R$ does not change any of this. ∎

Lemma 5 *If $q = d$ or $q = d+1$ and if the class $\{\varphi_n^q(s,t,\cdot,\cdot,\cdot) : s \in S_f, t \in T\}$ is a permissible, Euclidean class for a constant envelope, then $\sup\left|(P_n - P)\varphi_n^q\right| = o(n^{-1/2}\kappa_{zn}^{-1/2-q}\log n)$ almost surely $[P]$.*

Proof By change of variables

$$\begin{aligned}\sup P(\varphi_n^q)^2 &= \frac{1}{\kappa_{zn}^{1+2q}} \sup \iint f(s,y,m,x_{-1},t)^2 \\ &\qquad \times D^q K_z(u)^2 \zeta_2(\pi(s,t) + \kappa_{zn}u, y, m, x_{-1})\,du\,dP_2(y,m,x_{-1}) \\ &= O(\kappa_{zn}^{-1-2q}),\end{aligned}$$

and the conclusion of the lemma follows by theorem 2.37 of Pollard (1984). ∎

Lemma 6 *Define*

$$\tilde{\phi}^{d+1}(s,t) = \int f(s,y,m,x_{-1},t)x_{-1}D_1^{d+1}\zeta_2(\pi(s,t),y,m,x_{-1})\,dP_2(y,m,x_{-1}).$$

If $0 \le \alpha \le 1/2$, *then*

$$\sup\left|P_n\phi_n^d - P_n\varphi_n^d + (P_n\Pi)\phi^{d+1} - (P_n\Omega)'\tilde{\phi}^{d+1}\right| = o_p(n^{-\alpha}),$$

provided $n^{-1/2}\kappa_{zn}^{-3/2-d}\log n$ *is bounded and* $n^{\alpha-1}\kappa_{zn}^{-d-3} \to 0$.

Proof By the mean value theorem

$$\begin{aligned}\phi_n^d(s,t) = \varphi_n^d(s,t) + (\pi_n(s,t) - \pi(s,t) - P_n\Pi(s,t))\varphi_n^{d+1}(s,t) + (P_n\Pi(s,t))\varphi_n^{d+1}(s,t)\\ -(\beta_{n,-1} - \beta_{-1} - P_n\Omega)'\tilde{\varphi}_n^{d+1}(s,t) - (P_n\Omega)'\tilde{\varphi}_n^{d+1}(s,t) - \eta_n^{d+1}(s,t),\end{aligned}$$

where φ_n^d is defined in (58), $\tilde{\varphi}_n^{d+1}(s,t,m,x,y) = \varphi_n^{d+1}(s,t,m,x,y)x_{-1}$,

$$\begin{aligned}\eta_n^{d+1}(s,t,m,x,y) = f(s,y,m,x_{-1},t)\Big[(\beta_{n,-1} - \beta_{-1})'x_{-1} - (\pi_n(s,t) - \pi(s,t))\Big]\\ \times \frac{D^{d+1}K_z(\kappa_{zn}^{-1}(\beta_n'^{*}h - \pi_n^*(s,t))) - D^{d+1}K_z(\kappa_{zn}^{-1}(\beta'x - \pi(s,t)))}{\kappa_{zn}^{2+d}},\end{aligned}$$

$\beta_n{}^*$ is between β and β_n and $\pi_n^*(s,t)$ is between $\pi(s,t)$ and $\pi_n(s,t)$.

Now, $\{\varphi_n^{d+1}(s,t,\cdot,\cdot,\cdot) : s \in S_f, t \in T\}$ is a permissible and Euclidean class by lemma 4, so by lemmas 3 and 5 with $d+1$ replacing d

$$\sup\left|P_n\varphi_n^{d+1} - \phi^{d+1}\right| = o(n^{-1/2}\kappa_{zn}^{-3/2-d}\log n) + O(\kappa_{zn}^{k_z}) = o(1) \quad \text{a.s. } [P].$$

Since $\sup|\phi^{d+1}|$ is finite,

$$\sup\Big(|\pi_n - \pi - P_n\Pi||\varphi_n^{d+1}|\Big) = o_p(n^{-1/2}) \quad \text{a.s. } [P],$$

provided $n^{-1/2}\kappa_{zn}^{-3/2-d}\log n$ is bounded.

Similarly, $\sup|P_n\tilde{\varphi}_n^{d+1} - \tilde{\phi}^{d+1}| = o(1)$ almost surely $[P]$ and $\sup|\tilde{\phi}^{d+1}|$ is finite, so

$$|\beta_{n,-1} - \beta_{-1} - P_n\Omega|\sup|\tilde{\varphi}_n^{d+1}| = o_p(n^{-1/2}) \quad \text{a.s. } [P],$$

provided $n^{-1/2}\kappa_{zn}^{-3/2-d}\log n$ is bounded.

Now consider η_n^{d+1}. Define

$$\hat{\eta}_n^{d+1}(s,t,m,x,y) = |f(s,y,m,x_{-1},t)|(1+|x_{-1}|)^2\frac{c}{\kappa_{zn}^{3+d}},$$

where c is the constant from assumption 7. Then

$$\begin{aligned} |\eta_n^{d+1}(s,t,m,x,y)| &\leq |f(s,y,m,x_{-1},t)|\frac{1}{\kappa_{zn}^{3+d}}\Big(|\beta_{n,-1} - \beta_{-1}||x_{-1}| + |\pi_n(s,t) - \pi(s,t)|\Big)^2 \\ &\leq \Big(|\beta_{n,-1} - \beta_{-1}| + |\pi_n(s,t) - \pi(s,t)|\Big)^2 |\hat{\eta}_n^{d+1}(s,t,m,x,y)|. \end{aligned}$$

Change of variables to get

$$\begin{aligned} \sup P\hat{\eta}_n^{d+1} &= \frac{c}{\kappa_{zn}^{d+3}}\sup\iint\Big[|f(s,y,m,x_{-1},t)|(1+|x_{-1}|)^2\zeta_2(u,y,m,x_{-1})\Big] \\ &\qquad\qquad \times\, du\, dP_2(y,m,x_{-1}) \\ &= O(\kappa_{zn}^{-d-3}) \end{aligned}$$

and

$$\begin{aligned} \sup P(\hat{\eta}_n^{d+1})^2 &= \frac{c^2}{\kappa_{zn}^{6+2d}}\sup\iiint\Big[f(s,y,m,x_{-1},t)^2(1+|x_{-1}|)^4\zeta_2(u,y,m,x_{-1})\Big] \\ &\qquad\qquad \times\, du\, dP_2(y,m,x_{-1}) \\ &= O(\kappa_{zn}^{-6-2d}). \end{aligned}$$

The class $\{\hat{\eta}_n^{d+1}(s,t,\cdot,\cdot,\cdot) : s \in S_f, t \in T\}$ is permissible and Euclidean, because $|K_z|$ is of bounded variation when K_z is and $\{|f(s,\cdot,\cdot,\cdot,t)| : s \in S_f, t \in T\}$ is Euclidean when

$\{f(s,\cdot,\cdot,\cdot,t) : s \in S_f, t \in T\}$ is. Therefore,

$$\sup\big|(P_n - P)\hat{\eta}_n^{d+1}\big| = o\big(n^{-1/2}\kappa_{zn}^{-3-d}\log n\big) \quad \text{a.s. } [P]$$

by theorem 2.37 of Pollard (1984). Since $\sup_{st}\big[|\beta_{n,-1} - \beta_{-1}| + |\pi_n(s,t) - \pi(s,t)|\big]^2 = O_p(n^{-1})$ it follows that

$$\sup P_n\big|\eta_n^{d+1}\big| = O_p(n^{-1}\kappa_{zn}^{-d-3}) + o_p(n^{-3/2}\kappa_{zn}^{-3-d}\log n) \quad \text{a.s. } [P].$$

The right-hand side is $o_p(n^{-\alpha})$ because

$$n^{\alpha-3/2}\kappa_{zn}^{-3-d}\log n = (n^{-1/2}\kappa_{zn}^{-3/2-d}\log n)(n^{\alpha-1}\kappa_{zn}^{-d-3})\kappa_{zn}^{d+3/2}.$$

Combining these results gives the conclusions of the lemma. ∎

Lemma 7 *If* $0 \leq \alpha < 1/2$ *then* $\sup|P_n\phi_n^d - \phi^d| = o_p(n^{-\alpha})$, *provided* $n^{\alpha-1/2}\kappa_{zn}^{-1/2-d}\log n$ *and* $n^{-1/2}\kappa_{zn}^{-3/2-d}\log n$ *are bounded,* $n^{\alpha}\kappa_{zn}^{k_z} \to 0$ *and* $n^{\alpha-1}\kappa_{zn}^{-d-3} \to 0$.

Proof By lemma 6

$$\sup|P_n\phi_n^d - \phi^d| \leq \sup|P_n\varphi_n^d - \phi^d| + \sup|P_n\Pi|\sup|\phi^{d+1}| + |P_n\Omega|\sup|\tilde{\phi}^{d+1}| + o_p(n^{-\alpha}),$$

where $\tilde{\phi}^{d+1}$ is defined in lemma 6. By lemma 4 the class $\{\varphi_n^d(s,t,\cdot,\cdot,\cdot) : s \in S_f, t \in T\}$ is permissible and Euclidean, so by lemmas 3 and 5

$$\sup|P_n\varphi_n^d - \phi^d| = o(n^{-1/2}\kappa_{zn}^{-1/2-d}\log n) + O(\kappa_{zn}^{k_z}) \quad \text{a.s. } [P].$$

The conclusion of the lemma now follows by observing that $\sup|\phi^{d+1}|$ and $\sup|\tilde{\phi}^{d+1}|$ are finite and that $\sup|P_n\Pi|$ and $|P_n\Omega|$ are $O_p(n^{-1/2})$. ∎

We now turn to the leading terms. Again we consider first the case where β and π are known. Define for $(m,x,y,t) \in R^{r+2} \times T$

$$F_n^d(m,x,y,t) = \int_{S_f} \varphi_n^d(s,t,m,x,y)\,ds \quad \text{and} \quad F^d(m,x,y,t) = D_1^d a(\beta' x, y, m, x_{-1}, t),$$

where φ_n^d is defined in (58).

Lemma 8 $\sup\bigl|P(F_n^d - F^d)\bigr| = O(\kappa_{zn}^{k_z})$.

Proof By Fubini's theorem $PF_n^d(t) = \int_{S_f} P\varphi_n^d(s,t)\,ds$, and by integration by parts, change of variables and Fubini's theorem, $PF^d(t) = \int_{S_f} \phi^d(s,t)\,ds$. It follows that $P(F_n^d(t) - F^d(t)) = \int_{S_f} (P\varphi_n^d(s,t) - \phi^d(s,t))\,ds$. Now apply lemma 3. ∎

Lemma 9 *The classes $\{F_n^d(t) : t \in T\}$ and $\{F^d(t) : t \in T\}$ are Euclidean for constant envelopes, and $\sup\bigl|(P_n - P)(F_n^d(t) - F^d(t))\bigr| = o(\kappa_{zn}^{1/2} n^{-1/2} \log n)$ almost surely $[P]$ provided $\kappa_{zn}^{-1/2} n^{-1/2} \log n$ is nonincreasing.*

Proof By change of variables, definition of a and repeated integration by parts

$$F_n^d(m,x,y,t) = \int D_1^d a(s,y,m,x_{-1},t) \frac{K_z(\kappa_{zn}^{-1}(\beta' x - s))}{\kappa_{zn}^2}\,ds.$$

By another change of variables

$$F_n^d(m,x,y,t) - F^d(m,x,y,t) = \int [D_1^d a(\beta' x - \kappa_{zn} s, y, m, x_{-1}, t) - D_1^d a(\beta' x, y, m, x_{-1}, t)] K_z(s)\,ds \quad \text{a.s. } [P(m,x,y)]$$

and $\bigl|F_n^d(t) - F^d(t)\bigr| \le 2\eta_1 \nu$ almost surely $[P]$, where η_1 is a uniform almost sure $[P_2]$ bound on $D_1^d a(\beta' x, \cdot, \cdot, \cdot, t)$ and $\nu = \int \bigl|K_z(s)\bigr|\,ds$.

Use the conclusion of lemma 2.5 to cover the boundary of each $S_a(y,m,x_{-1},t)$ by c_0/κ_{zn}^{q-1} closed balls of radius κ_{zn} centered on the boundary of $S_a(y,m,x_{-1},t)$, and let $V_n(y,m,x_{-1},t)$ be the union of c_0/κ_{zn}^{q-1} closed balls with the same centers but radius $(1+\sqrt{q})\kappa_{zn}$. If $s \in S_a(y,m,x_{-1},t) - V_n(y,m,x_{-1},t)$ and $|s-u| < \sqrt{q}\kappa_{zn}$ then $u \in$

$S_a(y,m,x_{-1},t)$. Moreover, the volume of $V_n(y,m,x_{-1},t)$ is less than $c_0'\kappa_{zn}$ where $c_0' = c_0 2(1+\sqrt{q})^q \pi^{q/2}/q\Gamma(q/2)$. Define the function $w_n(m,x,y,t) = 1$ if $\beta' x \in V_n(y,m,x_{-1},t) \cap S_a(y,m,x_{-1},t)$ and $w_n(m,x,y,t) = 0$ else. Then

$$P|F_n^d(t) - F^d(t)| \leq P(w_n(t)|F_n^d(t) - F^d(t)|) + P((1-w_n(t))|F_n^d(t) - F^d(t)|).$$

By assumption there is a number η_2 bounding ζ_2, so

$$P(w_n(t)|F_n^d(t) - F^d(t)|) \leq 2\eta_1 \nu \eta_2 c_0' \kappa_{zn}.$$

For n so large that $\kappa_{zn} \leq q^{-1/2}$, if $\beta' x \in S_a(y,m,x_{-1},t) - V_n(y,m,x_{-1},t)$ and $s \in [-1,1]^q$ then $|\kappa_{zn} s| \leq \sqrt{q}\kappa_{zn} \leq 1$ and $\beta' x - \kappa_{zn} s \in S_a(y,m,x_{-1},t)$. Therefore, by lemma 2.5

$$P((1-w_n(t))|F_n^d(t) - F^d(t)|) \leq c_1 q^{1/2} \kappa_{zn} \nu.$$

Combining these results gives $\sup P|F_n^d(t) - F^d(t)| = O(\kappa_{zn})$ and

$$\sup P(F_n^d(t) - F^d(t))^2 \leq 2\eta_1 \nu \sup P|F_n^d(t) - F^d(t)| = O(\kappa_{zn}).$$

Since $F_n^d(t) - F^d(t)$ is bounded almost surely $[P]$ the conclusion of the lemma follows from theorem 2.37 of Pollard (1984) provided $\{F_n^d(t) - F^d(t) : t \in T\}$ is a Euclidean class. But $\{F^d(t) : t \in T\}$ is Euclidean by assumption, and $\{F_n^d(t) : t \in T\}$ is Euclidean by lemma 2.13 of Pakes and Pollard (1989), because

$$|F_n^d(m,x,y,t) - F_n^d(m,x,y,t^*)| \leq c_2 |t - t^*| \sup |K_z|$$

by lemma 2.6. By lemma 2.14 of Pakes and Pollard (1989) $\{F_n^d(t) - F^d(t) : t \in T\}$ is Euclidean. ∎

Lemma 10 *The class $\{\Phi^d(t) : t \in T\}$ is permissible and Euclidean for a constant envelope* *and* $\sup|P_n(\Phi_n^d - \Phi^d)| = o_p(n^{-1/2})$, *provided* $n^{1/2}\kappa_{zn}^{k_z} \to 0$, $n^{-1/2}\kappa_{zn}^{-3/2-d}\log n$ *is bounded,* $n^{-1/2}\kappa_{zn}^{-d-3} \to 0$, *and* $\kappa_{zn}^{-1/2} n^{-1/2}\log n$ *is nonincreasing.*

Proof As a class of functions of (m, x, y) indexed by $t \in T$ the term involving Ω in the definition of Φ^d is Euclidean by lemma 2.14(iv) of Pakes and Pollard (1989) because the integral is bounded over $t \in T$. The term involving Π is a Euclidean class indexed by $t \in T$ by lemma 2.13 of Pakes and Pollard (1989) and lemma 2.7. Since $\{D_1^d a(\cdot,\cdot,\cdot,\cdot,t) : t \in T\}$ is Euclidean by assumption and $\Phi^d(t)$ is uniformly bounded, $\{\Phi^d(t) : t \in T\}$ is Euclidean class for a constant envelope.

Now, by lemmas 6, 8 and 9

$$\begin{aligned}
&\sup|P_n(\Phi_n^d - \Phi^d)| \\
&\qquad \le \sup|P_n(F_n^d - F^d)| + c_s \sup|P_n\phi_n^d - P_n\varphi_n^d + \phi^{d+1}P_n\Pi - (\tilde{\phi}^{d+1})'P_n\Omega| \\
&\qquad = o_p(n^{-1/2})
\end{aligned}$$

where c_s is the volume of S_f. The lemma follows. ∎

A.3 Estimating Covariances

In this section we show how the covariance functions can be estimated. Let γ be the function from lemma 1 representing λ or ψ and define ν by $\nu(m,x,y,s,t) = \gamma(m,x,y,s)\gamma(m,x,y,t)$. The covariance function corresponding to γ is $P\nu$. Suppose we have an estimator γ_n of γ such that $\sup|P_n(\gamma_n - \gamma)| \to 0$ in probability. Then $P_n\nu_n$ is a uniformly consistent estimator of $P\nu$, because $\sup|P_n(\nu_n - \nu)| \to 0$ in probability (since γ is bounded) and $\sup|P_n\nu - P\nu| \to 0$ almost surely [P] by theorem II.24 of Pollard (1984). Therefore, to estimate the covariance functions all we need are uniformly consistent estimates of λ and ψ, that is, of the functions Φ_i, $i = 1, \ldots, 8$, defined in equations (36), (37), (45) and (46).

Let a_i be the a-function corresponding to f_i (see page 28). Define f_{in} and $D_1^d a_{in}$ by replacing β and every unknown function $(A, B, \Lambda, \ldots)$ in the formulae for f_i and $D_1^d a_i$ by its estimator $(\beta_n, A_n, B_n, \Lambda_n, \ldots)$; note that $D\Lambda(y) = -A(z,y)/B(z,y)D_1H(z,y)$ and

$D\Lambda^{-1}(y) = 1/D\Lambda(\Lambda^{-1}(y))$. An argument similar to those used to prove uniform convergence of the remainder terms R_n^Λ and R_n^Ψ shows that f_{in} and $D_1^d a_{in}$ are uniformly consistent estimators of the corresponding functions f_i and $D_1^d a_i$.

The function Π is estimated by 0 in the case of λ and by λ_n in the case of ψ. The function Ω depends on the particular choice of β_n and estimators are given in the references in the introduction. The densities ζ_1 and ζ_2 can be estimated consistently uniformly on S_D and I_Z by kernel estimators, see for example Bierens (1987).[10] The distributions P_1 and P_2 are simply estimated by their empirical distributions P_{1n} and P_{2n}.

Define Φ_{in} by replacing all unknown quantities in the formulae for Φ_i by the above estimators. The integrals in the definitions of Φ_{in} converge uniformly to their counterparts in the definitions of Φ_i because the integrands converge uniformly and have bounded supports and the expectations $P_{1n}(\cdot) - P_1(\cdot)$ and $P_{2n}(\cdot) - P_2(\cdot)$ converge uniformly by theorem II.24 of Pollard (1984). It follows that Φ_{in} are uniformly consistent estimators of the functions Φ_i.

[10]The article by Bierens is about kernel estimation of regression functions.

References

Bierens, H. J. (1987). Kernel estimators of regression functions. In T. F. Bewley (Ed.), *Advances in Econometrics*. New York: Cambridge University Press.

Breslow, N. and J. Crowley (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics 2*(3), 437–453.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics 35*, 303–316.

Härdle, W. and T. M. Stoker (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association 84*, 986–995.

Heckman, J. J. and B. Singer (1984, Mar). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52*, 271–320.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica 64*(1), 103–137.

Horowitz, J. L. and W. Härdle (1994). Direct semiparametric estimation of single-index models with discrete covariates. Working paper, University of Iowa, Department of Economics.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics 58*, 71–120.

Kiefer, N. M. (1988, Jun). Economic duration data and hazard functions. *Journal of Economic Literature 26*, 646–679.

Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *Annals of Statistics 12*, 766–774.

Murphy, S. A. (1991). Consistency in a proportional hazards model incorporating a random effect. Unpublished Manuscript, Department of Statistics, Pennsylvania State

University.

Murphy, S. A. (1992). Asymptotic theory for the frailty model. Technical Report No 108, Department of Statistics, Pennsylvania State University.

Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. I. A. Sørensen (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics 19*, 25–43.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *Annals of Statistics 15*(2), 780–799.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–1057.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989, Nov). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

Ridder, G. (1990, Apr). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies 57*, 167–182.