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

Author(s): Joel L. Horowitz
Source: *Econometrica*, Vol. 64, No. 1 (Jan., 1996), pp. 103-137
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/2171926
Accessed: 28-10-2017 08:52 UTC

Econometrica, Vol. 64, No. 1 (January, 1996), 103–137

SEMIPARAMETRIC ESTIMATION OF A REGRESSION MODEL WITH AN UNKNOWN TRANSFORMATION OF THE DEPENDENT VARIABLE

BY JOEL L. HOROWITZ[1]

This paper presents a method for estimating the model $\Lambda(Y) = \beta'X + U$, where Y is a scalar, Λ is an unknown increasing function, X is a vector of explanatory variables, β is a vector of unknown parameters, and U has unknown cumulative distribution function F. It is not assumed that Λ and F belong to known parametric families; they are estimated nonparametrically. This model generalizes a large number of widely used models that make stronger *a priori* assumptions about Λ and/or F. The paper develops $n^{1/2}$-consistent, asymptotically normal estimators of Λ, F, and quantiles of the conditional distribution of Y. Estimators of β that are $n^{1/2}$-consistent and asymptotically normal already exist. The results of Monte Carlo experiments indicate that the new estimators work reasonably well in samples of size 100.

KEYWORDS: Semiparametric estimation, transformation model, empirical process, unobserved heterogeneity.

1. INTRODUCTION

MANY MODELS THAT ARE WIDELY USED in applied econometrics and statistics have the form

$$\Lambda(Y) = \beta'X + U, \tag{1.1}$$

where Y is a scalar dependent variable, $\Lambda(\cdot)$ is an increasing function, X is a $q \times 1$ vector of observed explanatory variables, β is a $q \times 1$ vector of constant parameters, and U is an unobserved random variable that is independent of X. For example, the proportional hazards and accelerated failure time models have this form. Let F denote the cumulative distribution function (CDF) of U. This paper presents a method for estimating Λ and F in (1.1) when these functions and β are unknown. In contrast to existing methods for estimating Λ and F, it is not assumed that either function belongs to a known (finite-dimensional) parametric family. Both are estimated nonparametrically.

The focus of the paper is on developing $n^{1/2}$-consistent, asymptotically normal, nonparametric estimators of Λ and F. Estimates of Λ and F can be used to obtain predictions of Y conditional on X, estimate elasticities of Y with respect to components of X, and estimate the hazard function of Y conditional

[1] I thank Wolfgang Härdle, John Kennan, Charles Manski, Whitney Newey, James Powell, Geert Ridder, Gene Savin, Scott Thompson, Alexander Tsybakov, and the referees for helpful comments and suggestions. The research reported in this paper was supported in part by NSF Grants DMS-9208820 and SBR-9307677.

on X.[2] Estimators of β are already available. Equation (1.1) belongs to the class of single-index models considered by Han (1987a), Härdle and Stoker (1989), Ichimura (1993), Powell et al. (1989), and Sherman (1993). Methods described by these authors can be used to obtain $n^{1/2}$-consistent, asymptotically normal estimators of β. I make no attempt to develop a new estimator of β.

Equation (1.1) includes as special cases a large number of widely used and extensively investigated models that make stronger a priori assumptions than are made here about Λ and/or F. These models include:

(i) The log-linear regression and accelerated failure time models.

(ii) Transformation models in which Λ is specified up to a finite-dimensional parameter (e.g., Box and Cox (1964), Bickel and Doksum (1981), Johnson (1949), John and Draper (1980), and MacKinnon and Magee (1990)). Often in these models it is assumed that F is known up to a finite-dimensional parameter, but this is not necessary (see, e.g., Amemiya (1985, p. 250), and Han (1987b)). Newey (1990) and Robinson (1991) proposed asymptotically efficient estimators of the parameters of models in which Λ is parametric and F is nonparametric.

(iii) The parametric and semiparametric proportional hazards models. In these models $F(u) = 1 - \exp(-e^u)$, and $\exp[\Lambda(y)]$ is the integrated baseline hazard function. In the parametric proportional hazards model, Λ is known up to a finite-dimensional parameter. In the semiparametric proportional hazards model (Cox (1972)), Λ is unknown.[3]

(iv) The mixed proportional hazards (MPH) model or proportional hazards model with unobserved heterogeneity. In this model, $U = U_1 + U_2$, where U_1 has CDF $F(u) = 1 - \exp(-e^u)$ and U_2 is a random variable that represents unobserved heterogeneity. The integrated baseline hazard function is $\exp(\Lambda)$. Elbers and Ridder (1982), Heckman and Singer (1984), and Ridder (1990) have investigated the identifiability of the MPH model. Existing estimators for this model require either Λ or the distribution of U_2 to be specified up to a finite-dimensional parameter (Heckman and Singer (1984); Murphy (1991, 1992); Nielsen et al. (1992)). The estimator presented here removes this requirement.

Each of the foregoing models is a special case of (1.1) because each assumes that Λ and/or F belongs to a known finite-dimensional parametric family.[4] In this paper, neither Λ nor F is assumed to belong to such a family. It is not unusual for familiar parametric models to fit data poorly, and this has often been the main motivation for developing new parametric forms. By avoiding parametric assumptions about Λ and F, this paper avoids the need to search for

[2] Predictions, elasticities, and hazard functions also can be obtained nonparametrically from the nonparametric regression of Y on $b_n'X$, where b_n is a consistent estimator of β. But the resulting sampling errors are larger than with the methods described in this paper because the rate of convergence of nonparametric regression estimators is slower than $n^{-1/2}$.

[3] The proportional hazards model is usually written in the form $\lambda(Y|X) = \lambda_0(Y)\exp(-\beta'X)$, where $\lambda(\cdot|\cdot)$ is the conditional hazard function of Y and λ_0 is the "baseline" hazard function. It is easily shown that this is equivalent to (1.1) with $F(u) = 1 - \exp(-e^u)$.

[4] The definition of the mixed proportional hazards model does not require Λ or F to be specified parametrically but, as has been noted, existing estimators for the model impose this requirement.

suitable parametric forms as well as the possibility of inconsistent estimation and systematically erroneous prediction due to misspecification of Λ and/or F.

Denote the nonparametric estimators of Λ and F by Λ_n and F_n, respectively, where n is the size of the estimation sample. It will be shown that Λ_n and F_n are asymptotically normal in the sense that as $n \to \infty$, $n^{1/2}(\Lambda_n - \Lambda)$ and $n^{1/2}(F_n - F)$ converge in distribution to mean-zero Gaussian processes over suitable compact intervals of the real line. It will also be shown that Λ_n and F_n converge in probability to Λ and F uniformly over these intervals.[5] Finally, it will be shown how these results can be used to obtain predictions of Y conditional on X.

The main task of the paper is developing a $n^{1/2}$-consistent, asymptotically normal estimator of Λ. Given such an estimator, it is a relatively easy task to estimate F and predict Y. Breiman and Friedman (1985) and Hastie and Tibshirani (1990) proposed estimators of Λ for a somewhat more general transformation model, but these estimators are not $n^{1/2}$-consistent.

The remainder of the paper is organized as follows. Section 2 presents the estimators of Λ and F, derives their asymptotic distributional properties, and shows how they can be used to predict Y. Section 3 presents the results of a Monte Carlo investigation of the finite-sample properties of the estimators. Concluding comments are given in Section 4. The proofs of theorems are in Appendix A.

2. THE ESTIMATORS AND THEIR ASYMPTOTIC PROPERTIES

a. Motivation

It is useful to begin with an informal discussion that motivates the estimators of Λ and F.

Equation (1.1) continues to hold if Λ and U are replaced by $\Lambda + c$ and $U + c$, respectively, where c is any constant. Equation (1.1) also holds if Λ, β, and U are replaced by $c\Lambda$, $c\beta$, and cU for any $c > 0$. Therefore, location and scale normalizations are needed to make identification possible. My location normalization consists of setting $\Lambda(y_0) = 0$ for some finite y_0 that satisfies conditions given in Section 2c. With this location normalization, there is no centering assumption on U and no intercept term in X. To achieve scale normalization, I follow the literature on estimation of β in semiparametric single-index models and assume that X has at least one component whose β coefficient is nonzero and whose probability distribution conditional on the remaining components is absolutely continuous with respect to Lebesgue measure. Arrange the components of X so that the first component satisfies this condition, and let β_1 denote

[5] The convergence of Λ_n and F_n to Λ and F is in probability or almost surely according to whether the estimator of β converges in probability or almost surely. In this paper, I assume only that the estimator of β converges in probability.

the corresponding component of β. My scale normalization consists of setting $|\beta_1| = 1$.[6]

To begin the derivation of the estimator of Λ, observe that by (1.1), Y depends on X only through the index $Z \equiv \beta' X$. Let $G(\cdot|z)$ be the CDF of Y conditional on $Z = z$. Assume that Λ and F are differentiable and that G is differentiable with respect to both of its arguments. Define $\lambda(y) \equiv d\Lambda(y)/dy$, $f(y) \equiv dF(y)/dy$, $G_y(y|z) \equiv \partial G(y|z)/\partial y$, and $G_z(y|z) \equiv \partial G(y|z)/\partial z$.

Equation (1.1) implies that

$$G(y|z) = F[\Lambda(y) - z]. \tag{2.1}$$

Therefore, $G_y(y|z) = \lambda(y) f[\Lambda(y) - z]$, $G_z(y|z) = -f[\Lambda(y) - z]$, and $\lambda(y) = -G_y(y|z)/G_z(y|z)$ for any (y, z) such that $G_z(y|z) \neq 0$. It follows that

$$\Lambda(y) = -\int_{y_0}^{y} \left[G_y(v|z)/G_z(v|z)\right] dv \tag{2.2}$$

for any z such that the denominator of the integrand is nonzero over the range of integration.

Now let $w(\cdot)$ be a scalar-valued function on $\mathscr{R}$ with compact support S_w such that (a) the denominator of the integrand in (2.2) is nonzero for all $z \in S_w$ and $v \in [y_0, y]$, and (b)

$$\int_{S_w} w(z)\, dz = 1. \tag{2.3}$$

Then

$$\Lambda(y) = -\int_{y_0}^{y} \int_{S_w} w(z) \left[G_y(v|z)/G_z(v|z)\right] dz\, dv. \tag{2.4}$$

Equation (2.4) is the basis of the estimator of Λ proposed here. The unknown parameter β is replaced with a $n^{1/2}$-consistent estimator, b_n (e.g., one of the estimators listed in Section 1). The unknown function $G(y|z)$ is replaced with a nonparametric kernel estimator, $G_n(\cdot|z)$, of the CDF of Y conditional on $b_n' X = z$. G_z in (2.4) is replaced by $G_{nz} \equiv \partial G_n / \partial z$. G_y is replaced by a kernel estimator, G_{ny}, of the probability density function of Y conditional on $b_n' X = z$. The resulting estimator of Λ is

$$\Lambda_n(y) = -\int_{y_0}^{y} \int_{S_w} w(z) \left[G_{ny}(v|z)/G_{nz}(v|z)\right] dz\, dv. \tag{2.6}$$

As is well known, kernel estimators converge in probability at rates slower than $n^{-1/2}$, so $G_{ny}(v|z)/G_{nz}(v|z)$ is not $n^{1/2}$-consistent for $G_y(v|z)/G_z(v|z)$. However, integration over z and v in (2.5) creates an averaging effect that causes the integral and, therefore, Λ_n to converge at the rate $n^{-1/2}$. This is the reason for basing the estimator of Λ on (2.4) instead of (2.2). A similar

[6] The scale normalization implies that if (1.1) is a mixed proportional hazards model, $\exp(\Lambda)$ is a positive power of the integrated baseline hazard function. Estimating this power presents a difficult deconvolution problem that is not investigated here.

averaging effect occurs in semiparametric estimation of average derivatives (Härdle and Stoker (1989), Powell, et al. (1989)).[7]

I now turn to the estimator of F. If $\Lambda(y)$ were estimable over the entire interval $-\infty < y < \infty$, F could be estimated $n^{1/2}$-consistently by the empirical distribution function (EDF) of $U_n \equiv \Lambda_n(Y) - b_n'X$. However, because $\Lambda(y)$ may not be bounded as $y \to \pm\infty$ and the denominator of (2.4) must be kept away from 0, Λ can be estimated $n^{1/2}$-consistently only for y's in a compact interval $[y_2, y_1]$ that is a proper subset of the support of the distribution of Y. Therefore, U is estimated consistently only if the corresponding Y value is in $[y_2, y_1]$. Otherwise, since Λ is an increasing function, it is known that $U < \Lambda(y_2) - \beta'X$ or $U > \Lambda(y_1) - \beta'X$ according to whether $Y < y_2$ or $Y > y_1$, but the value of U is not estimated consistently. Because of this censoring-like phenomenon, the EDF of U_n is an inconsistent estimator of F.

To motivate the method for dealing with this problem, suppose for the moment that β is known and that $\Lambda(y)$ is known if $y \in [y_2, y_1]$. Since Λ is an increasing function, it is also known that $\Lambda(y) < \Lambda(y_2)$ if $y < y_2$ and $\Lambda(y) > \Lambda(y_1)$ if $y > y_1$. Set $U = \Lambda(Y) - Z$. Call the event $U \le u$ "observable" if it can be determined whether $U \le u$ for every value of Y. Note that the event $U \le u$ is observable if $\Lambda(y_2) - u < Z \le \Lambda(y_1) - u$. Since U and Z are independent, $F(u) = P[U \le u | \Lambda(y_2) - u < Z \le \Lambda(y_1) - u]$ whenever $P[\Lambda(y_2) - u < Z \le \Lambda(y_1) - u] > 0$. Moreover, the empirical probability that $U \le u$ conditional on the event $\Lambda(y_2) - u < Z \le \Lambda(y_1) - u$ estimates $F(u)$ consistently despite the censoring of U's corresponding to Y values outside of $[y_2, y_1]$. This suggests that $F(\cdot)$ can be estimated consistently by the EDF of U_n conditional on "estimated observability," that is on the event $\Lambda_n(y_2) - u < b_n'X \le \Lambda_n(y_1) - u$.

To make these ideas precise, let $p(\cdot)$ denote the probability density function of Z, and let $1(\cdot)$ be the indicator function. Then

$$F(u) = P[U \le u | \Lambda(y_2) - u < Z \le \Lambda(y_1) - u] = A(u)/B(u), \tag{2.6}$$

where

$$A(u) = E\{1(U \le u)1[\Lambda(y_2) - u < Z \le \Lambda(y_1) - u]\}, \tag{2.7}$$

[7] $\Lambda(y)$ can also be estimated as a sample average. Observe that $\Lambda(y) = -E\{w(Z)1(y_0 \le Y \le y)/[p(Z)G_z(Y|Z)]\}$, where $p(\cdot)$ is the probability density function of Z. Let $p_n(\cdot)$ be the estimator of $p(\cdot)$ that is given in equation (2.12), and define $Z_{ni} = b_n'X_i$. Then $\Lambda(y)$ can be estimated by the following sample analog of $-E\{w(Z)1(y_0 \le Y \le y)/[p(Z)G_z(Y|Z)]\}$:

$$-(1/n)\sum_{i=1}^{n} w(Z_{ni})1(y_0 \le Y_i \le y)/[p_n(Z_{ni})G_{nz}(Y_i|Z_{ni})].$$

It can be shown that this sample-average estimator is asymptotically equivalent to (2.5). However, in Monte Carlo experiments with $n = 100$, I have found that the mean-square error of the sample-average estimator is much larger than that of the estimator that is obtained by using quadrature to integrate (2.5). A referee has suggested replacing G_{ny}/G_{nz} with $-|G_{ny}/G_{nz}|$, thereby guaranteeing that $\Lambda_n(\cdot)$ is nondecreasing. This modification has no effect on the asymptotic distribution of Λ_n, but my experience in Monte Carlo experiments is that it produces a large bias in Λ_n when the sample size and data-generation process are such that Λ_n has decreasing segments with high probability.

and

$$B(u) = E\{1[\Lambda(y_2) - u < Z \le \Lambda(y_1) - u]\}. \tag{2.8}$$

The estimator of F is obtained by replacing unknown quantities with their sample analogs in (2.6)–(2.8). Specifically, let $\{Y_i, X_i: i = 1, \dots, n\}$ denote a random sample of (Y, X). Define $Z_{ni} \equiv b_n' X_i$ and $U_{ni} \equiv \Lambda_n(Y_i) - Z_{ni}$, where $\Lambda_n(y)$ is set equal to an arbitrarily large negative number if $y < y_2$ and an arbitrarily large positive number if $y > y_1$. The estimator is

$$F_n(u) = A_n(u)/B_n(u), \tag{2.9}$$

where

$$A_n(u) \equiv n^{-1} \sum_{i=1}^{n} 1(U_{ni} \le u) 1[\Lambda_n(y_2) - u < Z_{ni} \le \Lambda_n(y_1) - u] \tag{2.10}$$

and

$$B_n(u) \equiv n^{-1} \sum_{i=1}^{n} 1[\Lambda_n(y_2) - u < Z_{ni} \le \Lambda_n(y_1) - u]. \tag{2.11}$$

In Section 2d it is shown that F_n converges in probability to F uniformly over $u \in [u_0, u_1]$, where u_0 and u_1 are chosen so that $P[\Lambda(y_2) - u < Z \le \Lambda(y_1) - u] > 0$ whenever $u \in [u_0, u_1]$. In addition, $n^{1/2}(F_n - F)$ converges in distribution to a mean-zero Gaussian process.[8]

b. The Estimators of $G(y|z)$ and Its Derivatives

This section presents the estimators of $G(y|z)$ and its derivatives that are used in constructing Λ_n. Let K_Y and K_Z be kernel functions in the sense of nonparametric density estimation and regression. Let $\{h_{ny}\}$ and $\{h_{nz}\}$ $(n = 1, 2, \dots)$ be sequences of bandwidth parameters. Conditions that the kernel function and bandwidth parameters must satisfy are given in Section 2c. Among other things, to insure that certain bias and remainder terms in the asymptotic expansions of $n^{1/2}(\Lambda_n - \Lambda)$ and $n^{1/2}(F_n - F)$ vanish as $n \to \infty$, K_z (but not K_Y) must be a "higher-order" kernel, and h_{ny} and h_{nz} must converge to 0 faster than the asymptotically optimal rates in nonparametric density estimation and regression. A higher-order kernel is needed for K_Z because G_{nz} is a functional of derivatives of K_Z (see equation (2.14)). Derivative functionals converge relatively slowly, and the higher-order kernel for K_Z is needed to insure sufficiently rapid convergence. Higher-order kernels are used for similar reasons in average derivative estimation (Härdle and Stoker (1989), Powell et al. (1989)). Robinson (1988) used a higher-order kernel to reduce bias in semiparametric

[8] In some applications, such as estimating the conditional hazard function of Y, it is necessary to have a differentiable estimator of F. This can be accomplished by replacing $1(U_{ni} < u)$ in (2.10) with $Q[(u - U_{ni})/h_{nu}]$, where Q is a differentiable CDF and $h_{nu} \to 0$ at a suitable rate as $n \to \infty$. I do not pursue this elaboration here.

estimation of a semilinear regression model. Neither G_{nz} nor G_{ny} involves derivatives of K_Y, so K_Y need not be a higher-order kernel.

Recall that $Z_{ni} \equiv b_n' X_i$. Estimate $p(\cdot)$, the probability density function of Z, by

$$(2.12) \qquad p_n(z) = (nh_{nz})^{-1} \sum_{i=1}^{n} K_Z\left(\frac{Z_{ni} - z}{h_{nz}}\right).$$

The estimator of $G(y|z)$ is

$$(2.13) \qquad G_n(y|z) = [nh_{nz} p_n(z)]^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_Z\left(\frac{Z_{ni} - z}{h_{nz}}\right).$$

The estimator of $G_z(y|z)$ is obtained by differentiating $G_n(y|z)$:

$$(2.14) \qquad G_{nz}(y|z) = \partial G_n(y|z)/\partial z.$$

$G_y(y|z)$ is the probability density function of Y conditional on $Z = z$. It cannot be estimated by $\partial G_n(y|z)/\partial y$ because $G_n(y|z)$ is a step function of y. Instead, the following kernel density estimator is used:

$$(2.15) \qquad G_{ny}(y|z) = \left[nh_{ny}h_{nz}p_n(z)\right]^{-1} \sum_{i=1}^{n} K_Y\left(\frac{Y_i - y}{h_{ny}}\right) K_Z\left(\frac{Z_{ni} - z}{h_{nz}}\right).$$

Λ_n is obtained by substituting (2.14) and (2.15) into (2.5).

c. *Assumptions*

This section gives the assumptions used in proving $n^{1/2}$-consistency and asymptotic normality of Λ_n and F_n. These assumptions are stronger than needed to prove uniform convergence of Λ_n and F_n to Λ and F. Among other things, proving uniform convergence does not require the use of a higher-order kernel for K_Z or existence of certain higher-order derivatives associated with the use of a higher-order K_Z. However, the main objective of this paper is to achieve $n^{1/2}$-consistency and asymptotic normality, so the assumptions are presented in the form needed to accomplish this.

In what follows, $\tilde{X}$ is the $(q-1) \times 1$ vector consisting of components 2 through q of X, $\tilde{\beta}$ is the $(q-1) \times 1$ vector consisting of components 2 through q of β, β_1 is the first component of β, $Z \equiv \beta' X$, and $r \ge 6$ and $s \ge 2$ are integers.

I make the following assumptions about the data-generation process and the probability distribution of (Y, Z):

ASSUMPTION 1: *$\{Y_i, X_i$: $i = 1, \ldots, n\}$ is a random sample of (Y, X) in* (1.1).

ASSUMPTION 2: (a) *Either $\beta_1 = 1$ or $\beta_1 = -1$.* (b) *The distribution of the first component of X conditional on $\tilde{X} = \tilde{x}$ is absolutely continuous with respect to Lebesgue measure for every $\tilde{x}$ in the support of $\tilde{X}$.* (c) *$\tilde{X}$ has bounded support.*

Assumption 2 insures the existence of $p(\cdot)$ and $p(\cdot|\tilde{x})$, the probability density function of Z conditional on $\tilde{X}=\tilde{x}$. Assumption 2c can be relaxed at the expense of more complex proofs.

ASSUMPTION 3: (a) *U is independent of X.* (b) *The distribution of U is absolutely continuous with respect to Lebesgue measure.* (c) *Let $f(u) \equiv dF(u)/du$. There is an open subset, I_U, of the support of the distribution of U such that* $\sup\{f(u): u \in I_U\} < \infty$, $\inf\{f(u): u \in I_U\} > 0$ *and the derivatives $d^k f(u)/du^k$ $(k = 1, \ldots, r+s+1)$ exist and are uniformly bounded for $u \in I_U$.*

ASSUMPTION 4: *$\Lambda(\cdot)$ is a strictly increasing, differentiable function everywhere on the support of Y.*

ASSUMPTION 5: *There are open intervals of the real line, I_y and I_z, such that $y_0 \in I_y$ and:*

a. *$y \in I_y$ and $z \in I_z \Rightarrow \Lambda(y) - z \in I_U$.*

b. *$p(z)$ and $p(z|\tilde{x})$ are bounded uniformly over $z \in I_z$ and $\tilde{x}$ in the support of $\tilde{X}$. Moreover,* $\inf\{p(z): z \in I_z\} > 0$.

c. *The derivatives $d^k p(z)/dz^k$ and $\partial^k p(z|\tilde{x})/\partial z^k$ $(k = 1, \ldots, r+1)$ exist and are uniformly bounded for all $z \in I_z$ and $\tilde{x}$ in the support of $\tilde{X}$.*

d. *$\Lambda(y_0) = 0$, and* $\sup\{|\Lambda(y)|: y \in I_y\} < \infty$. *The derivatives $d^k\Lambda(y)/dy^k$ $(k = 1, \ldots, s+1)$ exist and are uniformly bounded over $y \in I_y$.*

The weight function $w(\cdot)$ is assumed to satisfy the following assumption.

ASSUMPTION 6: *S_w is compact, $S_w \subset I_z$,* (2.3) *holds, and $d^k w(z)/dw^k$ $(k = 1, \ldots, r+1)$ exists and is bounded for all $z \in I_z$.*

Assumptions 3–6 insure, among other things, that $G_y(y|z)$ and $G_z(y|z)$ exist and that the denominator in (2.4) is bounded away from 0.[9] The existence of higher-order derivatives of Λ, f, p, and w is needed to insure that bias terms associated with the kernel estimators G_{ny} and G_{nz} vanish sufficiently rapidly.

Let b_{n1} and $\tilde{b}_n$, respectively, denote the estimators of β_1 and $\tilde{\beta}$. Because β_1 and $b_{n1} = \pm 1$ by scale normalization, $b_{n1} = \beta_1$ with probability approaching 1 as $n \to \infty$. Therefore, for purposes of obtaining the asymptotic distributions of $n^{1/2}(\Lambda_n - \Lambda)$ and $n^{1/2}(F_n - F)$ it can be assumed that $b_{n1} = \beta_1$. The remaining components of b_n are assumed to satisfy the following assumption.

[9] Since the purpose of the weight function is to keep the denominator of (2.4) away from 0, it may seem that $w(z)$ can be replaced by the trimming function $1[G_{ny}(y|z) > c_n]$, $c_n \to 0$, as is done in some forms of average derivative estimation (Härdle and Stoker (1989)). Although I cannot rule out the possibility that such an approach can be made to work, my proof of $n^{1/2}$-consistency and asymptotic normality requires $w(\cdot)$ to be differentiable and independent of y. The proof does not work if $w(\cdot)$ is replaced by a trimming function.

ASSUMPTION 7: *There is a* $(q-1)\times 1$*-vector-valued function* $\Omega(y,x)$ *such that* $E\Omega(Y,X)=0$, *the components of* $E[\Omega(Y,X)\Omega(Y,X)']$ *are finite, and as* $n\to\infty$

$$n^{1/2}\left(\tilde{b}_n-\tilde{\beta}\right)=n^{-1/2}\sum_{i=1}^{n}\Omega(Y_i,X_i)+o_p(1).$$

This assumption is satisfied by all of the estimators of β mentioned in Section 1.

As was noted in Section 2b, the kernel K_Z must be higher order, and the bandwidths h_{ny} and h_{nz} must converge to 0 faster than the asymptotically optimal rates in nonparametric density estimation and regression. The precise conditions that must be satisfied by K_Y, K_Z and the bandwidths are as follows.

ASSUMPTION 8: K_Y *has support* $[-1,1]$, *is bounded and symmetrical about* 0, *has bounded variation, and satisfies*

$$\int_{-1}^{1} v^j K_Y(v)\,dv=\begin{cases}1 & \text{if } j=0,\\ 0 & \text{if } 1\le j\le s-1,\\ \text{nonzero} & \text{if } j=s.\end{cases}$$

K_Z *has support* $[-1,1]$, *is bounded and symmetrical about* 0, *and satisfies*

$$\int_{-1}^{1} v^j K_Z(v)\,dv=\begin{cases}1 & \text{if } j=0,\\ 0 & \text{if } 1\le j\le r-1,\\ \text{nonzero} & \text{if } j=r.\end{cases}$$

K_Z *is everywhere twice differentiable. The derivatives are bounded and have bounded variation. The second derivative satisfies* $|K_Z''(v_1)-K_Z''(v_2)|\le M|v_1-v_2|$ *for some* $M<\infty$.

Müller (1984) gives kernel functions satisfying these conditions.

ASSUMPTION 9: *As* $n\to\infty$, $nh_{nz}^{2r}\to 0$, $nh_{ny}^{2s}\to 0$, $1/(nh_{nz}^{8})\to 0$, *and* $(\log n)/(n^{1/2}h_{ny}^{1/2}h_{nz}^{3})\to 0$.

Assumptions 8 and 9 are satisfied, for example, if K_Y is a second-order kernel, K_Z is a sixth-order kernel, $h_{ny}\propto n^{-1/3}$, and $h_{nz}\propto n^{-1/10}$

d. Results

This section gives theorems establishing that Λ_n and F_n are uniformly consistent and that $n^{1/2}(\Lambda_n-\Lambda)$ and $n^{1/2}(F_n-F)$ converge in distribution to Gaussian processes.

To prove convergence in distribution of $n^{1/2}(\Lambda_n-\Lambda)$ and $n^{1/2}(F_n-F)$, it is necessary to embed these functions in a metric space with a suitable σ-field. I use the space of cadlag functions (right-continuous functions that have left limits) with the metric generated by the supremum norm and the σ-field generated by the closed balls in this metric. The following additional notation is used. Let $\Rightarrow$ denote convergence in distribution as $n \to \infty$, $\tilde{P}$ denote the CDF of $\tilde{X}$, and E_{YX} denote the expectation with respect to the joint distribution of (Y, X). Define $w'(z) = dw(z)/dz$, $p'(z) = dp(z)/dz$, $g_z(y, z) = G_z(y|z)p(z)$, $D(y, z) = \partial[w(z)/g_z(y, z)]/\partial z$, $\pi(u) = P[\Lambda(y_2) - u < Z \leq \Lambda(y_1) - u]$,

$$C_\Lambda = \int [w'(z)\tilde{x}p(z|\tilde{x})/p(z)]\, dz\, d\tilde{P}(\tilde{x}),$$

$$C_F(u) = \int \tilde{x}P[\Lambda(y_2) - u < Z \leq \Lambda(y_1) - u|\tilde{x}]\, d\tilde{P}(\tilde{x}),$$

$$\begin{aligned} J_y(Y, X) = {} & -w(Z)\left[\frac{1(y_0 < Y \leq y) - 1(y < Y \leq y_0)}{g_z(Y, Z)} + \frac{\Lambda(y)}{p(Z)}\right] \\ & + \int_{y_0}^{y}\left[\frac{w(Z)p'(Z)}{g_z(v, Z)p(Z)} + D(v, z)\right] \\ & \times \lambda(v)[1(Y \leq v) - G(v|Z)]\, dv - \Lambda(y)C_\Lambda'\Omega(Y, X), \end{aligned}$$

and

$$\begin{aligned} L_u(Y, X) = \pi(u)^{-1}\Bigg\{ & \{1[\Lambda(Y) \leq Z + u] - F(u)\} \\ & \times 1[\Lambda(y_2) - u < Z \leq \Lambda(y_1) - u] \\ & - f(u)\int_{y_2}^{y_1} J_y(Y, X)p[\Lambda(y) - u]\lambda(y)\, dy \\ & + f(u)C_F(u)'\Omega(Y, X)\Bigg\}. \end{aligned}$$

The following theorem establishes uniform consistency of Λ_n and convergence in distribution of $n^{1/2}(\Lambda_n - \Lambda)$.

THEOREM 1: *Let* $[y_2, y_1] \in I_y$. *Under Assumptions* 1–9:
a. $p\lim_{n \to \infty} \sup_{y_2 \leq y \leq y_1} |\Lambda_n(y) - \Lambda(y)| = 0$.
b. *For* $y \in [y_2, y_1]$, $n^{1/2}[\Lambda_n(y) - \Lambda(y)] \Rightarrow H_\Lambda(y)$, *where* $H_\Lambda(y)$ *is a tight Gaussian process with mean* 0 *and covariance function* $E[H_\Lambda(y)H_\Lambda(y')] = E_{YX}[J_y(Y, X)J_{y'}(Y, X)]$.

The covariance function of H_Λ can be estimated consistently by replacing unknown quantities with sample analogs. See Appendix B for details.

Theorem 1 is proved in 4 steps whose details are given in Appendix A:

1. Use Taylor series expansions to approximate G_{ny}/G_{nz} uniformly over $(y,z)\in I_y \times I_z$ by a linear functional of kernel estimators and $\Omega(Y,X)$. Show that the error made by the linear approximation is $o_p(n^{-1/2})$.
2. Use a uniform law of large numbers of Pollard (1984) to show that the quantity observed by replacing G_{ny}/G_{nz} in (2.5) with the linear approximation converges almost surely to Λ uniformly over $[y_2, y_1]$. This proves Theorem 1a.
3. Show that the quantity obtained by replacing G_{ny}/G_{nz} in (2.5) with the linear approximation is asymptotically equivalent to an empirical process after centering and normalization (i.e., it is asymptotically equivalent to a sum of the form $n^{-1/2}\sum_{i=1}^{n}[\Psi_y(Y_i, X_i) - E\Psi_y(Y,X)]$ for a suitable function $\Psi_y(\cdot,\cdot)$).
4. Use methods described by Pollard (1984) to prove convergence in distribution of the empirical process.

The next theorem establishes uniform consistency of F_n and convergence in distribution of $n^{1/2}(F_n - F)$.

THEOREM 2: *Let* $[u_0, u_1] \in I_U$ *and* $P[\Lambda(y_2) - u \le Z \le \Lambda(y_1) - u] > 0$ *whenever* $u \in [u_0, u_1]$. *Under Assumptions* 1–9:

a. $p\lim_{n\to\infty}\sup_{u_0\le u\le u_1}|F_n(u) - F(u)| = 0$,

b. *For* $u \in [u_0, u_1]$, $n^{1/2}[F_n(u) - F(u)] \Rightarrow H_F(u)$, *where* $H_F(u)$ *is a Gaussian process with mean* 0 *and covariance function* $E[H_F(u)H_F(u')] = E_{YX}[L_u(Y, X)L_{u'}(Y,X)]$.

The covariance function of H_F can be estimated consistently on (u_0, u_1) by replacing unknown quantities with sample analogs. The details are given in Appendix B.

If Λ and β were known, Theorem 2a could be proved by substituting Λ and β into (2.9)–(2.11) and applying the Glivenko-Cantelli Theorem. Since Λ and β are not known, it must also be shown that the error made by replacing Λ and β with Λ_n and b_n in (2.9)–(2.11) is uniformly $o_p(1)$. I obtain this result using methods described by Pollard (1984) and Pakes and Pollard (1989).

Theorem 2b is proved in 3 steps:

1. Use a Taylor series expansion to approximate $F_n - F$ by a linear functional of $A_n - A$ and $B_n - B$.
2. Use stochastic equicontinuity arguments to approximate $n^{1/2}(A_n - A)$ and $n^{1/2}(B_n - B)$ by empirical processes. Use this result and the result of step 1 to approximate $n^{1/2}(F_n - F)$ by an empirical process. Show that the approximation error is $o_p(1)$.
3. Use methods described by Pollard (1984) to prove convergence in distribution of the empirical process that approximates $n^{1/2}(F_n - F)$.

e. Predicting Y Conditional on X

The most familiar predictor of Y conditional on $X = x$ is a consistent estimator of $E(Y|X = x)$. This predictor cannot be used in the present setting, however, because $\Lambda(y)$ and $F(u)$ are not estimated consistently for all y and u.[10] An alternative is to use an estimator of the median or, possibly, some other quantile of the distribution of Y conditional on $X=x$. This section uses Λ_n and F_n to obtain estimators of quantiles of the conditional distribution of Y.

Let $u_\gamma (0<\gamma<1)$ denote the γ quantile of the distribution of U. The γ quantile of the distribution of Y conditional on $X=x$ is

$$y_\gamma(x) = \Lambda^{-1}(\beta' x + u_\gamma).$$

Let x be given. Assume that $u_\gamma \in [u_0, u_1]$ and $\beta' x \in [\Lambda(y_2) - u_\gamma, \Lambda(y_1) - u_\gamma]$. Define

$$u_{n\gamma} \equiv \inf\{u \in [u_0, u_1]\colon F_n(u) \geq \gamma\}$$

if this quantity exists and u_1 otherwise. Define

$$y_{n\gamma}(x) = \inf\{y \in [y_2, y_1]\colon \Lambda_n(y) \geq b_n' x + u_{n\gamma}\}$$

if this quantity exists and y_1 otherwise. The following theorem gives conditions under which $y_{n\gamma}(x)$ is uniformly consistent and $n^{1/2}[y_{n\gamma}(x) - y_\gamma(x)]$ converges in distribution.

THEOREM 3: *Given any $\epsilon > 0$ and $\gamma \in (0,1)$, define $S_{\epsilon\gamma} = \{x\colon y_\gamma(x) \in [y_2+\epsilon, y_1 - \epsilon]\}$. Let $u_\gamma \in [u_0+\epsilon, u_1-\epsilon]$. Let Assumptions* 1–9 *hold. Then*

a. $p \lim_{n\to\infty} \sup_{x \in S_{\epsilon\gamma}} |y_{n\gamma}(x) - y_\gamma(x)| = 0$.

b. *For $x \in S_{\epsilon\gamma}$, $n^{1/2}[y_{n\gamma}(x) - y_\gamma(x)] \Rightarrow -H_\gamma(x)$, where H_γ is a mean-*0 *Gaussian process whose covariance function is the same as the covariance function of*

$$\lambda[y_\gamma(x)]^{-1}\left\{\Omega(Y,X)'\tilde{x} - L_{u_\gamma}(Y,X) f(u_\gamma)^{-1} - J_{y_\gamma(x)}(Y,X)\right\}.$$

A consistent estimator of the covariance function is given in Appendix B.

3. MONTE CARLO EXPERIMENTS

A small set of Monte Carlo experiments was carried out to investigate the finite-sample properties of the estimators of Λ, F, and $y_{0.50}(x) = \text{med}(Y|X=x)$. Samples were generated by simulation from (1.1) using three different Λ's. With additive constants reflecting their location normalizations, these are:

Experiment 1: $\Lambda(y) = y + 2$.
Experiment 2: $\Lambda(y) = \log y + 2$.
Experiment 3: $\Lambda(y) = (1/13)\sinh(2y) + 2.0992$.

X was a scalar with the $N(0,1)$ distribution in all of the experiments, and $U \sim N(\mu, 1)$, where $\mu = 2$ in experiments 1 and 2, and $\mu = 2.0992$ in experiment 3. Because X is a scalar, β is determined by scale normalization and was not

[10] Of course, $E(Y|X=x)$ can be estimated consistently by nonparametric regression, but the nonparametric regression estimator's rate of convergence in probability is less than the rate of $n^{-1/2}$ obtained here.

estimated. This focuses the experiments on the estimators of Λ, F, and $y_{0.50}$, which are the objects of interest in this paper.

The kernel and weight functions used in the experiments are

$$K_Y(x) = (15/16)(1-x^2)^2 1(|x| \leq 1), \tag{3.1}$$
$$K_Z(x) = (315/2048)(15 - 140x^2 + 378x^4 - 396x^6 + 143x^8)1(|x| \leq 1),$$

and $w(x) = 0.5 \cdot 1(|x| \leq 1)$. K_Y is a second-order kernel and K_Z is sixth order. Both are taken from Müller (1984). The weight function $w(\cdot)$ does not satisfy the differentiability requirement of Assumption 6. This is not important in a finite sample because with probability 1 the discontinuities of w are contained in intervals in which there are no observations of X.

I used bandwidths that were found through Monte Carlo experimentation to roughly minimize the integrated empirical mean square error of $y_{n,0.50}(x)$ over the interval $-2 \leq x \leq 2$. The bandwidths are shown in Table I. In experiments 1 and 3, $y_2 = y_0 = -2$, and $y_1 = 2$. In experiment 2, $y_2 = y_0 = e^{-2}$, and $y_1 = e^2$.

The sample size in all experiments was $n = 100$. Larger samples were not used because of the very long computing times they entail. There were 100 replications per experiment. The experiments were carried out in GAUSS, and GAUSS random number generators were used.

The results are summarized graphically in Figures 1–3 for experiments 1–3, respectively. The left-hand panels of the figures show the means of 100 estimates of Λ, F, and $y_{0.50}$ (solid lines) and the true values of these functions (dotted lines). The right-hand panels show the estimates of Λ, F, and $y_{0.50}$ obtained from the first 5 replications of each experiment (solid lines) and the true values of these functions (dotted lines). The piecewise-linear appearance of some of the estimates is an artifact of the plotting procedure, which consisted of connecting 20 equally spaced points with straight lines. It can be seen that the true functions and the means of the estimates are quite close to one another. The individual estimates are variable, not surprisingly, but most are reasonable approximations to the functions they estimate.

To provide an indication of relative finite-sample efficiencies of the estimator presented in this paper and other familiar estimators, I computed the integrated empirical mean square errors (IEMSE's) of the new estimator of $y_{0.50}(x)$ and the estimators of $y_{0.50}(x)$ obtained from:

1. Maximum likelihood estimation of (1.1) under the assumption that Λ and F are known but β is not. In this case, the estimator of $y_{0.50}(x)$ is $\Lambda^{-1}(b_n' x)$, where b_n is the maximum likelihood estimator of β.

TABLE I

BANDWIDTHS USED IN THE EXPERIMENTS

	New Estimator		Nonparametric Regression
$\Lambda(y)$	h_Y	h_Z	h
y	2.0	3.25	1.0
$\log(y)$	1.0	3.25	1.5
$(1/13)\sinh(2y)$	0.5	3.75	1.0

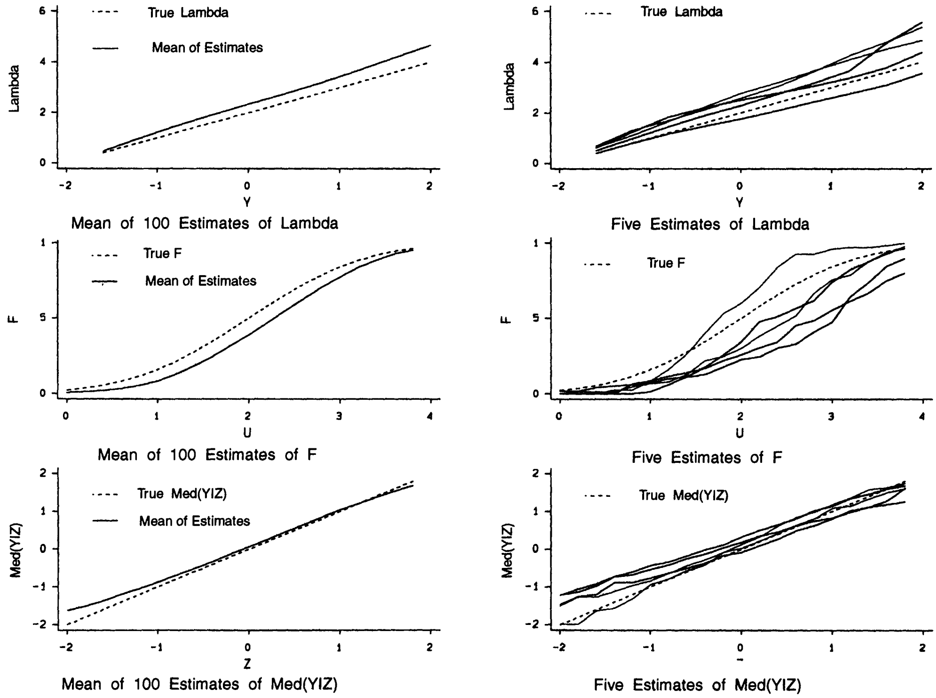

FIGURE 1.—Estimation results for the linear model.

2. Median nonparametric regression of Y on X. In this case, the estimator of $y_{0.50}(x)$ is the median of the distribution corresponding to the following nonparametric estimate of the conditional CDF of Y:

$$F_n(y|X=x) = [nhp_n(x)]^{-1} \sum_{i=1}^{n} 1(Y_i \leq y) K_Y\left(\frac{X_i - x}{h}\right),$$

where K_Y is the second-order kernel (3.1), h is a bandwidth parameter, and

$$p_n(x) = (nh)^{-1} \sum_{i=1}^{n} K_Y\left(\frac{X_i - x}{h}\right).$$

For each experiment, bandwidth h was chosen through Monte Carlo experimentation to minimize the IEMSE of the estimator of $y_{0.50}$ from the median nonparametric regression. The resulting values of h are shown in Table I. In all experiments, the interval of integration of the IEMSE was $-2 \leq x \leq 2$.

The IEMSE's for the various estimators and experiments are shown in Table II. Not surprisingly, the IEMSE of the maximum likelihood estimator is smaller than those of both the new and the nonparametric regression estimators. The new estimator has a smaller IEMSE than the nonparametric estimator in two of

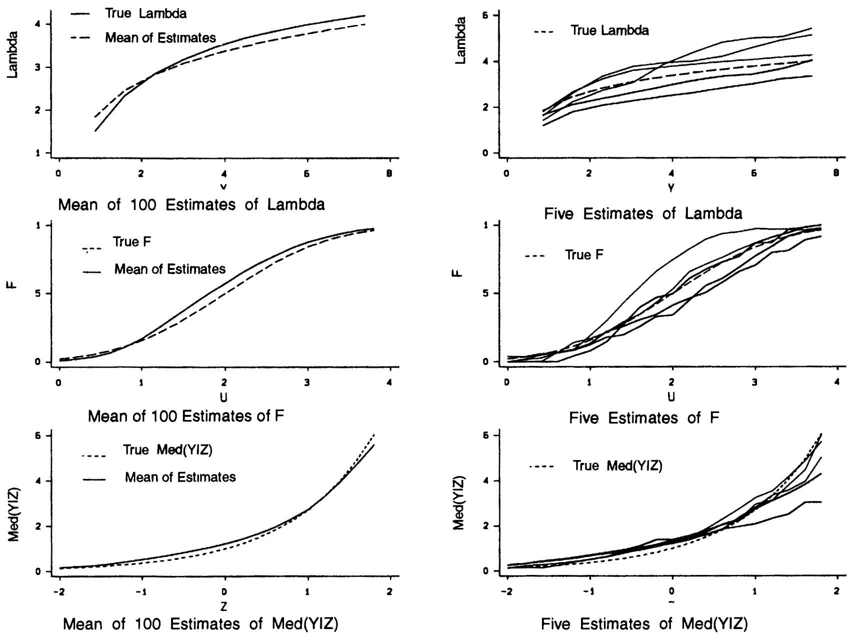

FIGURE 2.—Estimation results for the loglinear model.

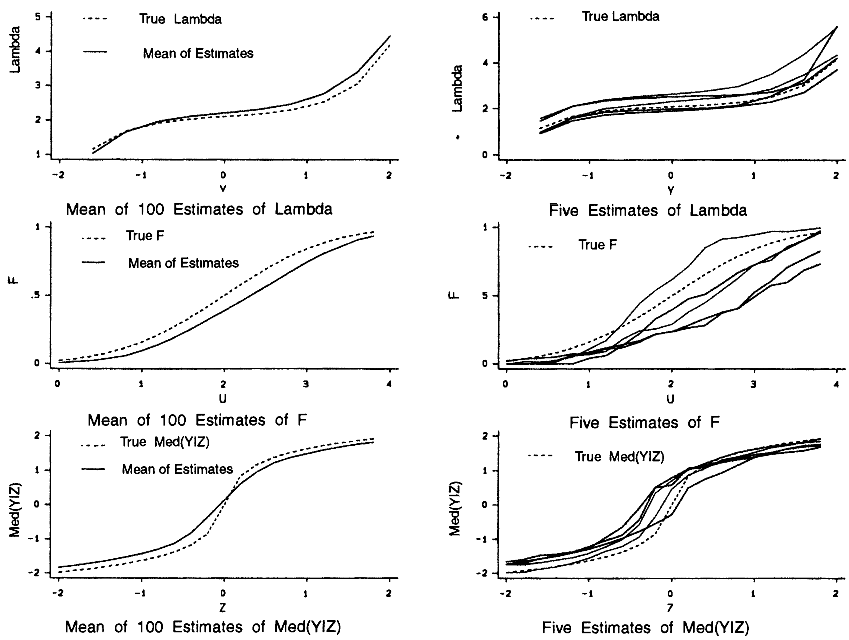

FIGURE 3.—Estimation results for the sinh model.

TABLE II

EMPIRICAL INTEGRATED MEAN SQUARE ERRORS OF ESTIMATORS OF MED($Y|Z$)

$\Lambda(y)$	OLS	Nonparametric Regression	New Estimator
y	0.046	0.104	0.065
$\log(y)$	0.401	0.603	0.517
$(1/13)\sinh(2y)$	0.072	0.122	0.131

the three experiments. In the third experiment, the IEMSE of the new estimator is slightly larger than that of the nonparametric estimator.

The theory presented in Section 2 insures that the new estimator converges in probability faster than the nonparametric estimator, whose rate of convergence is $n^{-2/5}$, but the theory provides no information about the relative efficiencies of the two estimators in finite samples. It is not unusual in semiparametric estimation for large samples to be needed to obtain the efficiency benefits promised by asymptotic theory. Therefore, it is not too surprising that, depending on the details of the model, samples of size 100 or more may be needed to make the IEMSE of the new estimator less than that of the nonparametric estimator.

4. CONCLUSIONS

This paper has presented a method for estimating a regression model with an unknown transformation of the dependent variable and an unknown distribution of the random error term. All of the unknown functions are estimated nonparametrically. The estimators are $n^{1/2}$ consistent and, after centering and normalization, converge in distribution to mean-zero Gaussian processes. Models that can be estimated with the new method include the semiparametric proportional hazards model with unobserved heterogeneity whose probability distribution is unknown.

There are several ways in which the methods presented in this paper can be extended. One is to permit censoring of the dependent variable, as often happens with duration data. Another is to permit weakly dependent data. Both of these extensions can be carried out using methods similar to the ones used here, but the details differ sufficiently from those here to require separate treatment. Finally, there is the problem of bandwidth selection. Presumably, asymptotically optimal bandwidths can be calculated from higher-order asymptotic expansions of the estimators developed here. Once the asymptotically optimal bandwidths are found, it should be possible to use a plug-in procedure to obtain bandwidths in applications.

Dept. of Economics, University of Iowa, Iowa City, IA 52242, U.S.A.

Manuscript received March, 1993; final revision received December, 1994.

APPENDIX A: PROOFS OF THEOREMS

Define $g_y(y,z) \equiv G_y(y|z)p(z)$, $g_z(y,z) \equiv G_z(y|z)p(z)$, $g_{ny}(y,z) \equiv G_{ny}(y|z)p_n(z)$, and $g_{nz}(y,z) \equiv G_{nz}(y|z)p_n(z)$. Equation (2.5) holds if G_{ny} and G_{nz} are replaced with g_{ny} and g_{nz}. For proving theorems, it is more convenient to use g_{ny} and g_{nz} than G_{ny} and G_{nz}. Define a Euclidean class of functions as in Pakes and Pollard (1989, p. 1032). It is assumed throughout that Assumptions 1–9 of the text hold.

Proofs are given for the case $y_0 = y_2$. To obtain proofs for other location normalizations, let $*$ denote quantities corresponding to $y_0 = y^* \neq y_2$, and let quantities without this symbol correspond to $y_0 = y_2$. The proof of Theorem 1 for $y_0 = y^*$ is obtained from the proof given here by observing that $\Lambda^*(y) = \Lambda(y) - \Lambda(y^*)$ and $\Lambda_n^*(y) = \Lambda_n(y) - \Lambda_n(y^*)$. The proofs of Theorems 2 and 3 for $y_0 = y^*$ are obtained from the proofs of these theorems given here by replacing X with $X^* = (X', 1)'$, β with $\beta^* = (\beta', \Lambda(y^*))'$, and b_n with $b_n^* = (b_n', \Lambda_n(y^*))'$.

Lemmas 1–5 establish asymptotic forms and uniform rates of convergence of g_{ny} and g_{nz}. These are used in proving Lemma 6, which obtains the linear approximation to G_{ny}/G_{nz} that is used in proving Theorem 1. Lemma 7 shows that substituting the linear approximation into (2.5) yields a quantity that, when multiplied by $n^{1/2}$, is asymptotically equivalent to an empirical process. Lemma 8 establishes a property of this process that is used to prove Theorem 1.

Because the arguments used to prove Lemmas 1–5 are essentially the same, only Lemma 1 is proved in detail. Lemmas 2–5 are stated without proof.

LEMMA 1: *Define*

$$g_{ny}^{(1)}(y,z) \equiv (nh_{ny}h_{nz})^{-1} \sum_{i=1}^{n} K_Y\left(\frac{Y_i - y}{h_{ny}}\right) K_Z\left(\frac{Z_i - z}{h_{nz}}\right),$$

$$\Gamma_1(y,z) \equiv -\int \tilde{x} \frac{\partial}{\partial z}[G_y(y|z)p(z|\tilde{x})]\, d\tilde{P}(\tilde{x}),$$

and

$$S_\Omega = n^{-1} \sum_{i=1}^{n} \Omega(Y_i, X_i).$$

As $n \to \infty$

$$|g_{ny}(y,z) - g_y(y,z) - \Gamma_1(y,z)' S_\Omega| = o_p\left[(\log n)/(nh_{ny}h_{nz})^{1/2}\right]$$

and

$$g_{ny}(y,z) = g_{ny}^{(1)}(y,z) + \Gamma_1(y,z)' S_\Omega + o_p(n^{-1/2})$$

uniformly over $(y,z) \in [y_0, y_1] \times S_w$.

PROOF: By the mean value theorem of differential calculus

$$g_{ny}(y,z) = g_{ny}^{(1)}(y,z) + g_{ny}^{(2)}(y,z)'(\tilde{b}_n - \tilde{\beta}) \tag{A1}$$

where

$$g_{ny}^{(2)}(y,z) \equiv (nh_{ny}h_{nz}^2)^{-1} \sum_{i=1}^{n} K_Y\left(\frac{Y_i - y}{h_{ny}}\right) K_Z'\left(\frac{\beta_n^{*\prime} X_i - z}{h_{nz}}\right) \tilde{X}_i,$$

$K_Z'(v) \equiv dK_Z(v)/dv$, and β_n^* is between b_n and β.

1. By example 2.10 of Pakes and Pollard (1989, p. 1033), the class of functions of ζ indexed by (h,z) of the form $Q_{hz}(\zeta)=K_Z[(\zeta-z)/h]$ $(z\in\mathscr{R}, h>0)$ is Euclidean. The class of functions of ψ of the form $Q_{hy}(\psi)=K_Y[(\psi-y)/h]$ $(y\in\mathscr{R}, h>0)$ also is Euclidean. Therefore, by Lemma 2.14 of Pakes and Pollard (1989, p. 1035), the class of functions of the form $Q_{shyz}(\psi,\zeta)=K_Y[(\psi-y)/h_y]K_Z[(\zeta-z)/h_z]$ $(y,z\in\mathscr{R}; h_y,h_z>0)$ is Euclidean. In addition,

$$E\left[K_Y\left(\frac{Y-y}{h_{ny}}\right)K_Z\left(\frac{Z-z}{h_{nz}}\right)\right]^2$$

$$=\int\left[K_Y\left(\frac{\psi-y}{h_{ny}}\right)K_Z\left(\frac{\zeta-z}{h_{nz}}\right)\right]^2 G_y(\psi|\zeta)p(\zeta)\,d\psi\,d\zeta$$

$$=h_{ny}h_{nz}\int K_Y(\psi)^2K_Z(\zeta)^2G_y(h_{ny}\psi+y|h_{nz}\zeta+z)$$

$$\times p(h_{nz}\zeta+z)\,d\psi\,d\zeta\le Mh_{ny}h_{nz}$$

for some finite M. By Theorem 2.37 of Pollard (1984, p. 34)

$$\text{(A2)}\qquad \sup_{y,z}|g_{ny}^{(1)}(y,z)-Eg_{ny}^{(1)}(y,z)|=o\left[(\log n)/(nh_{ny}h_{nz})^{1/2}\right]$$

as $n\to\infty$ almost surely. In addition,

$$\text{(A3)}\qquad E[g_{ny}^{(1)}(y,z)-g_y(y,z)]$$

$$=(h_{ny}h_{nz})^{-1}\int K_Y\left(\frac{\psi-y}{h_{ny}}\right)K_Z\left(\frac{\zeta-z}{h_{nz}}\right)G_y(\psi|\zeta)p(\zeta)\,d\psi\,d\zeta-g_y(y,z)$$

$$=\int K_Y(\psi)K_Z(\zeta)G_y(h_{ny}\psi+y|h_{nz}\zeta+z)p(h_{nz}\zeta+z)\,d\psi\,d\zeta-g_y(y,z).$$

Expand the integrand of (A3) in a Taylor series about $h_{ny}=h_{nz}=0$ through order s for h_{ny} and r for h_{nz} and use Assumption 8 to obtain

$$\text{(A4)}\qquad E[g_{ny}^{(1)}(y,z)-g_y(y,z)]=O(h_{ny}^s)+O(h_{nz}^r)$$

uniformly over $(y,z)\in[y_0,y_1]\times S_w$. By Assumption 9, h_{ny}^s and h_{nz}^r are $o[(\log n)/(nh_{ny}h_{nz})^{1/2}]$. Therefore, combining (A2) and (A4) yields

$$\text{(A5)}\qquad |g_{ny}^{(1)}(y,z)-g_y(y,z)|=o\left[(\log n)/(nh_{ny}h_{nz})^{1/2}\right]$$

almost surely uniformly over $(y,z)\in[y_0,y_1]\times S_w$.

2. Convergence of $g_{ny}^{(2)}$: $g_{ny}^{(2)}(y,z)=g_{ny}^{(3)}(y,z)+R_n(y,z)$, where

$$g_{ny}^{(3)}(y,z)\equiv(nh_{ny}h_{nz}^2)^{-1}\sum_{i=1}^{n}K_Y\left(\frac{Y_i-y}{h_{ny}}\right)K_Z'\left(\frac{Z_i-z}{h_{nz}}\right)\tilde{X}_i$$

and

$$R_n(y,z) = (nh_{ny}h_{nz}^2)^{-1} \sum_{i=1}^{n} K_Y\left(\frac{Y_i - y}{h_{ny}}\right)\left[K_Z'\left(\frac{\beta_n^{*\prime}X_i - z}{h_{nz}}\right) - K_Z'\left(\frac{Z_i - z}{h_{nz}}\right)\right]\tilde{X}_i.$$

Consider $R_n(y,z)$. Assumption 8 implies that K_Z' satisfies a Lipschitz condition, so for some $M < \infty$

$$\text{(A6)} \qquad \|R_n(y,z)\| \le M(nh_{ny}h_{nz}^3)^{-1}\|\beta_n^* - \beta\| \sum_{i=1}^{n}\left|K_Y\left(\frac{Y_i - y}{h_{ny}}\right)\right|\|\tilde{X}_i\|^2$$

$$\equiv R_n^*(y)\|\beta_n^* - \beta\|,$$

where $\|\cdot\|$ denotes the Euclidean norm. As in Lemma 1, the class of functions $K_Y[(\psi - y)/h]$ is Euclidean, and the class $Q(x) = \|\tilde{x}\|$ is Euclidean because it consists of a single function. Therefore, by Lemma 2.14 of Pakes and Pollard (1989), the summand of R_n^* is Euclidean. Application of Theorem 2.37 of Pollard (1984) yields

$$\text{(A7)} \qquad \sup_y |R_n^*(y) - ER_n^*(y)| = o[(\log n)/(n^{1/2}h_{ny}^{1/2}h_{nz}^3)]$$

almost surely. In addition

$$ER_n^*(y) = M(h_{ny}h_{nz}^3)^{-1}\int\left|K_Y\left(\frac{\psi - y}{h_{ny}}\right)\right|\|\tilde{x}\|^2 G_y(\psi|\zeta)p(\zeta|\tilde{x})\,d\psi\,d\zeta\,d\tilde{P}(\tilde{x})$$

$$= Mh_{nz}^{-3}\int|K_Y(\psi)|\,\|\tilde{x}\|^2 G_y(h_{ny}\psi + y|\zeta)p(\zeta|\tilde{x})\,d\psi\,d\zeta\,d\tilde{P}(\tilde{x})$$

$$\text{(A8)} \qquad = O(h_{nz}^{-3})$$

uniformly over $y \in I_y$. Since $\|\beta_n^* - \beta\| = O_p(n^{-1/2})$ by Assumption 7, combining (A6)–(A8) yields

$$\text{(A9)} \qquad R_n(y,z) = O_p\left[(nh_{nz}^6)^{-1/2}\right]$$

uniformly over $y \in [y_0, y_1]$ and all z.

Now consider $g_{ny}^{(3)}$. K_Z' has bounded variation by Assumption 8, so by Example 2.10 and Lemma 2.14 of Pakes and Pollard (1989), the summand in $g_{ny}^{(3)}$ forms a Euclidean class of functions. Application of Theorem 2.37 of Pollard (1984) yields

$$\text{(A10)} \qquad \sup_{y,z}\|g_{ny}^{(3)}(y,z) - Eg_{ny}^{(3)}(y,z)\| = o[(\log n)/(n^{1/2}h_{ny}^{1/2}h_{nz}^{3/2})]$$

almost surely. In addition,

$$Eg_{ny}^{(3)}(y,z)$$

$$= (h_{ny}h_{nz}^2)^{-1}\int K_Y\left(\frac{\psi - y}{h_{ny}}\right)K_Z'\left(\frac{\zeta - z}{h_{nz}}\right)\tilde{x}G_y(\psi|\zeta)p(\zeta|\tilde{x})\,d\psi\,d\zeta\,d\tilde{P}(\tilde{x})$$

$$\text{(A11)} \qquad = h_{nz}^{-1}\int K_Y(\psi)K_Z'(\zeta)\tilde{x}G_y(h_{ny}\psi + y|h_{nz}\zeta + z)p(h_{nz}\zeta + z|\tilde{x})\,d\psi\,d\zeta\,d\tilde{P}(\tilde{x}).$$

Expand the integrand of (A11) in a Taylor series about $h_{ny} = h_{nz} = 0$ through order s for h_{ny} and r for h_{nz} and use Assumption 8 to obtain

$$\text{(A12)} \qquad Eg_{ny}^{(3)}(y,z) = \Gamma_1(y,z) + O(h_{ny}^s) + O(h_{nz}^r)$$

uniformly over $(y,z) \in [y_0, y_1] \times S_w$. Combining (A10) and (A12) yields

$$\text{(A13)} \qquad |g_{ny}^{(3)}(y,z) - \Gamma_1(y,z)| = o[(\log n)/(n^{1/2}h_{ny}^{1/2}h_{nz}^{3/2})]$$

almost surely uniformly over $(y,z) \in [y_0, y_1] \times S_w$. The lemma follows by combining (A1), (A5), (A9), and (A13) with Assumption 7 and $\tilde{b}_n - \tilde{\beta} = O_p(n^{-1/2})$. *Q.E.D.*

LEMMA 2: *Define*

$$g_{nz}^{(1)}(y,z) \equiv -(nh_{nz}^2)^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_Z'\left(\frac{b_n' X_i - z}{h_{nz}}\right),$$

$$g_z^{(1)}(y,z) \equiv \frac{\partial}{\partial z}[G(y|z)p(z)],$$

$$g_{nz}^{(a)}(y,z) \equiv -(nh_{nz}^2)^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_Z'\left(\frac{Z_i - z}{h_{nz}}\right),$$

and

$$\Gamma_2(y,z) \equiv -\int \frac{\partial^2}{\partial z^2}[G(y|z)p(z|\tilde{x})]\tilde{x}\, d\tilde{P}(\tilde{x}).$$

As $n \to \infty$

$$|g_{nz}^{(1)}(y,z) - g_z^{(1)}(y,z) - \Gamma_2(y,z)' S_\Omega| = o_p\left[(\log n)/(nh_{nz}^3)^{1/2}\right]$$

and

$$g_{nz}^{(1)}(y,z) = g_{nz}^{(a)}(y,z) + \Gamma_2(y,z)' S_\Omega + o_p(n^{-1/2})$$

uniformly over $(y,z) \in [y_0, y_1] \times S_w$.

LEMMA 3: *Define*

$$p_n^{(1)}(z) = (nh_{nz})^{-1} \sum_{i=1}^{n} K_Z\left(\frac{Z_i - z}{h_{nz}}\right),$$

and

$$\Gamma_3(z) \equiv -\int \tilde{x} p'(z|\tilde{x})\, d\tilde{P}(\tilde{x}),$$

where $p'(z|\tilde{x}) \equiv \partial^2 p(z|\tilde{x})/\partial z^2$. *As* $n \to \infty$

$$|p_n(z) - p(z) - \Gamma_3(z)' S_\Omega| = o_p[(\log n)/(nh_{nz})^{1/2}]$$

and

$$p_n(z) = p_n^{(1)}(z) + \Gamma_3(z)' S_\Omega + o_p(n^{-1/2})$$

uniformly over $z \in S_w$.

LEMMA 4: *Define* $p_n'(z) \equiv dp_n(z)/dz$,

$$p_n'^{(1)}(z) = -(nh_{nz}^2)^{-1} \sum_{i=1}^{n} K_Z'\left(\frac{Z_i - z}{h_{nz}}\right),$$

and

$$\Gamma_4(z) \equiv -\int \tilde{x} p''(z|\tilde{x})\, d\tilde{P}(\tilde{x}),$$

where $p''(z|\tilde{x}) \equiv \partial^2 p(z|\tilde{x})/\partial z^2$. *As* $n \to \infty$

$$|p_n'(z) - p'(z) - \Gamma_4(z)' S_\Omega| = o_p\left[(\log n)/(nh_{nz}^3)^{1/2}\right]$$

and

$$p_n'(z) = p_n'^{(1)}(z) + \Gamma_4(z)' S_\Omega + o_p(n^{-1/2})$$

uniformly over $z \in S_w$.

LEMMA 5: *Define*

$$g_{nz}^{(2)}(y, z) \equiv (nh_{nz})^{-1} \sum_{i=1}^{n} 1(Y_i \leq y) K_Z\left(\frac{b_n' X_i - z}{h_{nz}}\right),$$

$$g_{nz}^{(d)}(y, z) \equiv (nh_{nz})^{-1} \sum_{i=1}^{n} 1(Y_i \leq y) K_Z\left(\frac{Z_i - z}{h_{nz}}\right),$$

and

$$\Gamma_5(y, z) \equiv -\int \tilde{x} \frac{\partial}{\partial z}[G(y|z) p(z|\tilde{x})]\, d\tilde{P}(\tilde{x}).$$

As $n \to \infty$

$$|g_{nz}^{(2)}(y, z) - G(y|z)p(z) - \Gamma_5(y, z)' S_\Omega| = o_p[(\log n)/(nh_{nz})^{1/2}]$$

and

$$g_{nz}^{(2)}(y, z) = g_{nz}^{(d)}(y, z) + \Gamma_5(y, z)' S_\Omega + o_p(n^{-1/2})$$

uniformly over $(y, z) \in [y_0, y_1] \times S_W$.

LEMMA 6: *Define*

$$Q_n(y,z) \equiv [nh_{nz}g_z(y,z)]^{-1}\sum_{i-1}^{n}\left\{h_{ny}^{-1}K_Y\left(\frac{Y_i-y}{h_{ny}}\right)\right.$$
$$-\lambda(y)[1(Y_i \le y) - G(y|z)]$$
$$\left.\cdot[p'(z)/p(z)]\right\}K_Z\left(\frac{Z_i-z}{h_{nz}}\right)$$
$$-\lambda(y)[nh_{nz}^2 g_z(y,z)]^{-1}\sum_{i=1}^{n}[1(Y_i \le y) - G(y|z)]K_Z'\left(\frac{Z_i-z}{h_{nz}}\right)$$

and

$$\Gamma_0(y,z) \equiv -\lambda(y)\frac{d}{dz}\left\{p(z)^{-1}\int \tilde{x}p(z|\tilde{x})\,d\tilde{P}(\tilde{x})\right\}.$$

As $n \to \infty$,

$$[G_{ny}(y|z)/G_{nz}(y|z) - G_y(y|z)/G_z(y|z)] = Q_n(y,z) + \Gamma_0(y,z)'S_\Omega + o_p(n^{-1/2})$$

uniformly over $(y,z) \in [y_0, y_1] \times S_w$.

PROOF: $G_{ny}/G_{nz} = g_{ny}/g_{nz}$. Therefore, a Taylor series expansion yields

$$G_{ny}(y|z)/G_{nz}(y|z) - G_y(y|z)/G_z(y|z) \tag{A14}$$
$$= g_z(y,z)^{-1}[g_{ny}(y,z) - g_y(y,z)]$$
$$-\left[g_y(y,z)/g_z(y,z)^2\right][g_{nz}(y,z) - g_z(y,z)]$$
$$+O\{[g_{nz}(y,z) - g_z(y,z)]^2\}$$
$$+O\{[g_{ny}(y,z) - g_y(y,z)][g_{nz}(y,z) - g_z(y,z)]\}.$$

In addition, a Taylor series expansion of $g_{nz}(y,z) - g_y(y,z)$ yields

$$g_{nz}(y,z) - g_z(y,z) = g_{nz}^{(1)}(y,z) - g_z(y,z) - \frac{g_{nz}^{(2)}(y,z)p_{nz}'(z)}{p(z)}$$
$$\cdot\left[1 - \frac{p_{nz}(z) - p(z)}{p(z)}\right] + O\{[p_{nz}(z) - p(z)]^2\}.$$

The lemma follows by substituting this result and the conclusions of Lemmas 1–5 into (A14) and making use of $\lambda(y) = -g_y(y,z)/g_z(y,z)$, $G_y(y|z) = -\lambda(y)G_z(y|z)$, and $\partial G_y(y|z)/\partial z = -\lambda(y)\partial^2 G(y|z)/\partial z^2$. *Q.E.D.*

LEMMA 7: *Define* $J_y(Y,Z)$ *as in the text. As* $n \to \infty$

$$\Lambda_n(y) - \Lambda(y) = n^{-1}\sum_{i=1}^{n} J_y(Y_i, Z_i) + o_p(n^{-1/2})$$

uniformly over $y \in [y_0, y_1]$.

PROOF: By (2.4), (2.5), and Lemma 6

$$\Lambda_n(y) - \Lambda(y) = -\int_{y_0}^{y}\int_{-\infty}^{\infty} w(z) Q_n(v,z)\,dz\,dv$$
$$-\int_{y_0}^{y}\int_{-\infty}^{\infty} w(z)\Gamma_0(v,z)'\,dv\,dz\,S_\Omega + o_p(n^{-1/2})$$

uniformly over $y \in [y_0, y_1]$.

1. Integrating over z: Write $Q_n(y,z) = Q_{n1}(y,z) + Q_{n2}(y,z)$, where

$$Q_{n1}(y,z) \equiv [nh_{nz}g(y,z)]^{-1}\sum_{i=1}^{n}\left\{h_{ny}^{-1}K_Y\left(\frac{Y_i - y}{h_{ny}}\right) - \lambda(y)[1(Y_i \le y) - G(y|z)]\right.$$
$$\left.\cdot[p'(z)/p(z)]\right\}K_Z\left(\frac{Z_i - z}{h_{nz}}\right)$$

and

$$Q_{n2}(y,z) \equiv -\lambda(y)[nh_{nz}^2 g_z(y,z)]^{-1}\sum_{i=1}^{n}[1(Y_i \le y) - G(y|z)]K_Z'\left(\frac{Z_i - z}{h_{nz}}\right).$$

Substitute $\zeta = (z - Z_i)/h_{nz}$ in each term of Q_{n1} and use the symmetry of K_Z to obtain

$$\int_{-\infty}^{\infty} w(z) Q_{n1}(y,z)\,dz$$
$$= n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{w(h_{nz}\zeta + Z_i)}{g_z(y, h_{nz}\zeta + Z_i)}\left\{h_{ny}^{-1}K_Y\left(\frac{Y_i - y}{h_{ny}}\right)\right.$$
$$\left.-\lambda(y)[1(Y_i \le y) - G(y|h_{nz}\zeta + Z_i)]\frac{p'(h_{nz}\zeta + Z_i)}{p(h_{nz}\zeta + Z_i)}\right\}K_Z(\zeta)\,d\zeta.$$

Expand the integrand in a Taylor series through order r about $h_{nz} = 0$ and make use of the fact that moments 1 through $r-1$ of K_Z are 0 to obtain

$$\text{(A15)}\qquad \int_{-\infty}^{\infty} w(z) Q_{n1}(y,z)\,dz = n^{-1}\sum_{i=1}^{n}\frac{w(Z_i)}{g_z(y, Z_i)}\left\{h_{ny}^{-1}K_Y\left(\frac{Y_i - y}{h_{ny}}\right) - \lambda(y)[1(Y_i \le y)\right.$$
$$\left.-G(y|Z_i)]\frac{p'(Z_i)}{p(Z_i)}\right\} + O(h_{nz}^r)$$

uniformly over $y \in [y_0, y_1]$. Now substitute $\zeta = (z - Z_i)/h_{nz}$ in each term of Q_{n2} and use the symmetry of K_Z to obtain

$$\int_{-\infty}^{\infty} w(z) Q_{n2}(y,z)\,dz = \lambda(y)(nh_{nz})^{-1}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{w(h_{nz}\zeta + Z_i)}{g_z(y, h_{nz}\zeta + Z_i)}[1(Y_i \le y)$$
$$-G(y|h_{nz}\zeta + Z_i)]K_Z'(\zeta)\,d\zeta.$$

Expand the integrand in a Taylor series through order $r+1$ about $h_{nz}=0$ and observe that moments 0 and 2 through r of K_Z' are zero to obtain

$$\text{(A16)} \qquad \int_{-\infty}^{\infty} w(z)Q_{n2}(y,z)\,dz = -\lambda(y)n^{-1}\sum_{i=1}^{n}\{D(y,Z_i)[1(Y_i\le y)-G(y|Z_i)] - w(Z_i)/p(Z_i)\} + O(h_{nz}^r)$$

uniformly over $y\in[y_0,y_1]$. Finally, use integration by parts to obtain

$$\text{(A17)} \qquad \int_{-\infty}^{\infty} w(z)\Gamma_0(y,z)\,dz = \lambda(y)\int_{-\infty}^{\infty}[w'(z)\tilde{x}p(z|\tilde{x})/p(z)]\,dz d\tilde{P}(\tilde{x}).$$

2. Integrating over y: Define

$$Q_{n1}^{(a)}(y) \equiv -n^{-1}\sum_{i=1}^{n} w(Z_i)\int_{y_0}^{y}[g_z(v,Z_i)h_{ny}]^{-1}K_Y\left(\frac{Y_i-v}{h_{ny}}\right)dv.$$

Combining (A15)–(A17) and integrating over y yields

$$\begin{aligned}\text{(A18)} \qquad \Lambda_n(y)-\Lambda(y) = {} & Q_{n1}^{(a)}(y) - [\Lambda(y)/n]\sum_{i=1}^{n} w(Z_i)/p(Z_i)\\ & + n^{-1}\sum_{i=1}^{n}\int_{y_0}^{y}\lambda(v)\left[\frac{w(Z_i)p'(Z_i)}{g_z(v,Z_i)p(Z_i)} + D(v,Z_i)\right]\\ & \times[1(Y_i\le y)-G(v|Z_i)]\,dv\\ & - \Lambda(y)C_\Lambda' S_\Omega + o_p(n^{-1/2})\end{aligned}$$

uniformly over $y\in[y_0,y_1]$. Now consider $Q_{n1}^{(a)}$. Let $E_{Y|Z}$ denote the expectation over $Y_1,\dots,Y_n$ conditional on $Z_1,\dots Z_n$.

$$\begin{aligned}\text{(A19)} \qquad & E_{Y|Z}Q_{n1}^{(a)}(y)\\ & = -n^{-1}\sum_{i=1}^{n} w(Z_i)\int_{y_0}^{y}\int_{-\infty}^{\infty}[g_z(v,Z_i)h_{ny}]^{-1}K_Y\left(\frac{\psi-v}{h_{ny}}\right)G_y(\psi|Z_i)\,d\psi\,dv\\ & = -n^{-1}\sum_{i=1}^{n} w(Z_i)\int_{y_0}^{y}\int_{-1}^{1} g_z(v,Z_i)^{-1}K_Y(\psi)G_y(h_{ny}\psi+v|Z_i)\,d\psi\,dv.\end{aligned}$$

Expand the integrand of (A19) in a Taylor series through order s about $h_{ny}=0$ and use the fact that $G_y(v|Z_i)/g_z(v,Z_i) = -\lambda(v)/p(Z_i)$ to obtain

$$E_{Y|Z}Q_{n1}^{(a)}(y) = n^{-1}\sum_{i=1}^{n} w(Z_i)\Lambda(y)/p(Z_i) + O(h_{ny}^s)$$

uniformly over $y\in[y_0,y_1]$ and $Z_1,\dots,Z_n$. In addition,

$$E_{Y|Z}n^{-1}\sum_{i=1}^{n} w(Z_i)1(y_0\le Y_i\le y)/g_z(Y_i,Z_i) = -n^{-1}\sum_{i=1}^{n} w(Z_i)\Lambda(y)/p(Z_i).$$

Therefore, uniformly over $y\in[y_0,y_1]$ and $Z_1,\dots,Z_n$

$$\text{(A20)} \qquad E_{Y|Z}\left\{Q_{n1}^{(a)}(y) + n^{-1}\sum_{i=1}^{n} w(Z_i)1(y_0\le Y_i\le y)/g_z(Y_i,Z_i)\right\} = O(h_{ny}^s).$$

Define

$$R_n(y) \equiv Q_{n1}^{(a)}(y) + n^{-1} \sum_{i=1}^{n} w(Z_i) 1(y_0 \le Y_i \le y)/g_z(Y_i, Z_i)$$

(A21)
$$= (nh_{ny})^{-1} \sum_{i=1}^{n} w(Z_i) \left\{ -\int_{y_0}^{y} g_z(v, Z_i)^{-1} K_Y\left(\frac{Y_i - v}{h_{ny}}\right) dv + h_{ny} 1(y_0 \le Y_i \le y)/g_z(Y_i, Z_i) \right\}.$$

It is not difficult to show that the summand in (A21) satisfies the conditions of Theorem 2.37 of Pollard (1984) and that as a consequence of this theorem

(A22)
$$\sup_{y \in [y_0, y_1]} |R_n(y) - ER_n(y)| = o[h_{ny}^{1/2}(\log n)/n^{1/2}]$$

almost surely. Combining (A20) and (A22) yields $R_n(y) = o_p(n^{-1/2})$ uniformly over $y \in [y_0, y_1]$. Therefore, in (A18) replacing $Q_{n1}^{(a)}$ with

$$-n^{-1} \sum_{i=1}^{n} w(Z_i) 1(y_0 \le Y_i \le y)/g_z(Y_i, Z_i)$$

causes an error whose size is $o_p(n^{-1/2})$ uniformly over $y \in [y_0, y_1]$. *Q.E.D.*

LEMMA 8: *Consider $J_y(Y, Z)$ ($y \in [y_0, y_1]$) to form a class of functions of (Y, X) indexed by y. This class is Euclidean.*

PROOF: The term $-w(Z)1(y_0 \le Y \le y)/g_z(Y, Z)$ forms a Euclidean class by Lemma 2.14 of Pakes and Pollard (1989) because the indicator function is a Euclidean class and $w(Z)/g_z(Y, Z)$ is trivially Euclidean because it is a single function. The remaining terms may be proved to be Euclidean by using Lemma 2.13 of Pakes and Pollard, and the sum of all the terms is Euclidean by a further application of Lemma 2.14. *Q.E.D.*

PROOF OF THEOREM 1: *Part a*: By Lemma 7, it suffices to prove uniform convergence to 0 of

$$n^{-1} \sum_{i=1}^{n} J_{y1}(Y_i, X_i) - \Lambda(y) C_\Lambda' n^{-1} \sum_{i=1}^{n} \Omega(Y_i, X_i),$$

where $J_{y1}(Y, X) = J_y(Y, X) + \Lambda(y) C_\Lambda' \Omega(Y, X)$. Almost sure uniform convergence of the first term follows from Lemma 8 and Theorem 2.24 of Pollard (1984). Almost sure uniform convergence of the second term follows from boundedness of Λ for $y \in [y_0, y_1]$ and the strong law of large numbers.

Part b: By Lemma 7, it suffices to prove convergence in distribution of

$$\Delta_n(y) \equiv n^{-1/2} \sum_{i=1}^{n} J_y(Y_i, X_i).$$

It is easily shown that $E_{YX}\Delta_n = 0$. By the Lindeberg-Levy Theorem and the Cramer-Wold device, the finite-dimensional distributions of Δ_n are asymptotically multivariate normal with covariance matrix $E_{YX}[J_y(Y,X)J_{y'}(Y,X)]$. By Theorem 7.13 of Pollard (1984) there is a Gaussian random element, Δ, of the space of cadlag functions on $[y_0, y_1]$ whose finite-dimensional distributions have this covariance matrix and whose sample paths are bounded and uniformly continuous. Δ is tight by Theorem 1.4 of Billingsley (1968). By Lemma 8 and Lemma 2.16 of Pakes and Pollard (1989), Δ_n satisfies the following stochastic equicontinuity condition: for each $\eta > 0$ and $\epsilon > 0$ there is a $\delta > 0$ such that

$$\lim_{n\to\infty} \sup P\left[\sup_{[\delta]} |\Delta_n(y) - \Delta_n(y')| > \eta\right] < \epsilon,$$

where $[\delta] \equiv \{(y,y')\colon E[J_y(Y,X) - J_{y'}(Y,X)]^2 < \delta^2\}$. $\Delta_n \Rightarrow \Delta$ now follows from Theorem 5.3 of Pollard (1984). *Q.E.D.*

PROOF OF THEOREM 2a: Choose any $\epsilon > 0$. By Theorem 1a, the following inequalities hold uniformly over $u \in [u_0, u_1]$ with probability approaching 1 as $n \to \infty$:

$$\text{(A23)} \qquad n^{-1}\sum_{i=1}^{n} 1(U_i \le u - \epsilon)1[-u + \epsilon < Z_i \le \Lambda(y_1) - u - \epsilon]$$
$$\le A_n \le n^{-1}\sum_{i=1}^{n} 1(U_i \le u + \epsilon)1[-u - \epsilon < Z_i \le \Lambda(y_1) - u + \epsilon]$$

and

$$\text{(A24)} \qquad n^{-1}\sum_{i=1}^{n} 1[-u + \epsilon < Z_i \le \Lambda(y_1) - u - \epsilon]$$
$$\le B_n \le n^{-1}\sum_{i=1}^{n} 1[-u - \epsilon < Z_i \le \Lambda(y_1) - u + \epsilon].$$

The outer terms of (A23) and (A24) converge uniformly and almost surely by the Glivenko-Cantelli Theorem. Therefore,

$$F(u-\epsilon)P[-u+\epsilon < Z \le \Lambda(y_1) - u - \epsilon] - \epsilon$$
$$\le A_n(u) \le F(u+\epsilon)P(-u-\epsilon < Z \le \Lambda(y_1) - u + \epsilon] + \epsilon$$

and

$$P[-u+\epsilon < Z \le \Lambda(y_1) - u - \epsilon] - \epsilon$$
$$\le A_n(u) \le P(-u-\epsilon < Z \le \Lambda(y_1) - u + \epsilon] + \epsilon$$

uniformly over $u \in [u_0, u_1]$ with probability approaching 1 as $n \to \infty$. Since F and the distribution of Z are absolutely continuous and ϵ is arbitrary, $A_n(u) \to^p A(u)$ and $B_n(u) \to^p B(u)$ uniformly over $u \in [u_0, u_1]$. The theorem now follows from that fact that $F(u) = A(u)/B(u)$. *Q.E.D.*

The proof of Theorem 2b uses Lemma 9 below and the following notation. For any (possibly random) function $h(y,x)$, define $\tilde{E}(h) = \int h(y,x)\,dP(y,x)$, where P is the CDF of (Y,X). $\tilde{E}$ is not the expectation operator because it does not average over any sources of randomness in h. For $u \in [u_0, u_1]$ define

$$E_n(u) \equiv n^{-1/2} \sum_{i=1}^{n} \{1[\Lambda_n(Y_i) - b_n' X_i \le u] 1[-u < b_n' X_i \le \Lambda_n(y_1) - u]$$

$$-\tilde{E} 1[\Lambda_n(y) - b_n' x \le u] 1[-u < b_n' x \le \Lambda_n(y_1) - u]\},$$

$$E_n^*(u) \equiv n^{-1/2} \sum_{i=1}^{n} \{1[\Lambda(Y_i) - \beta' X_i \le u] 1[-u < \beta' X_i \le \Lambda(y_1) - u]$$

$$-E 1[\Lambda(Y) - \beta' X \le u] 1[-u < \beta' X \le \Lambda(y_1) - u]\};$$

$$e_n(u) \equiv n^{-1/2} \sum_{i=1}^{n} \{1[-u < b_n' X_i \le \Lambda_n(y_1) - u]$$

$$-\tilde{E} 1[-u < b_n' x \le \Lambda_n(y_1) - u]\},$$

and

$$e_n^*(u) \equiv n^{-1/2} \sum_{i=1}^{n} \{1[-u < \beta' X_i \le \Lambda(y_1) - u] - E 1[-u < \beta' X \le \Lambda(y) - u]\}.$$

LEMMA 9:
a. $p \lim_{n \to \infty} \sup_{u \in [u_0, u_1]} |E_n(u) - E_n^*(u)| = 0.$
b. $p \lim_{n \to \infty} \sup_{u \in [u_0, u_1]} |e_n(u) - e_n^*(u)| = 0.$

PROOF: Only part a is proved here. The proof of part b is similar. Define $H_{n\Lambda}(y) \equiv n^{1/2}[\Lambda_n(y) - \Lambda(y)]$, and define $H_\Lambda(y)$ as in the text. Also, define $H_{n\beta} \equiv n^{1/2}(\tilde{b}_n - \tilde{\beta})$, and let H_β denote the limit in distribution of $H_{n\beta}$. Then

$$E_n(u) = n^{-1/2} \sum_{i=1}^{n} \{1[\Lambda(Y_i) - Z_i + n^{-1/2} H_{n\Lambda}(Y_i) - n^{-1/2} H_{n\beta}' \tilde{X}_i \le u]$$

$$\cdot 1[-u < Z_i + n^{-1/2} H_{n\beta}' \tilde{X}_i \le \Lambda(y_1) + n^{-1/2} H_{n\Lambda}(y_1) - u]$$

$$-\tilde{E} 1[\Lambda(y) - z + n^{-1/2} H_{n\Lambda}(y) - n^{-1/2} H_{n\beta}' x \le u]$$

$$\cdot 1[-u < z + n^{-1/2} H_{n\beta}' \tilde{x} \le \Lambda(y_1) + n^{-1/2} H_{n\Lambda}(y_1) - u]\}.$$

Let $\epsilon > 0$. By Theorem 1, there are compact sets $K_{\Lambda\epsilon}^*$ and $K_{\beta\epsilon}$ such that $P[(H_\Lambda, H_\beta) \in K_{\Lambda\epsilon}^* \times K_{\beta\epsilon}] > 1 - \epsilon/2$. By using Lemma 7.9 of Pollard (1984), it can be shown that there are a set $K_{\Lambda\epsilon} \subset K_{\Lambda\epsilon}^*$ and constants c and d such that $P[(H_\Lambda, H_\beta) \in K_{\Lambda\epsilon} \times K_{\beta\epsilon}] > 1 - \epsilon$ and

$$\text{(A25)} \qquad |H_\Lambda(y) - H_\Lambda(y')| \le c|y - y'|^d$$

on $K_{\Lambda\epsilon}$ for all sufficiently small $|y - y'|$. It can be assumed without loss of generality that $K_{\Lambda\epsilon}$ is closed and, therefore, compact. Let $\{H_{\Lambda\delta i}: i = 1, \ldots, I_{\Lambda\delta}\}$ and $\{H_{\beta\delta i}: i = 1, \ldots, I_{\beta\delta}\}$ be δ-nets for $K_{\Lambda\epsilon}$

and $K_{\beta\epsilon}$, respectively. Let $E_{njk}(u)$ denote the quantity obtained from $E_n(u)$ by replacing $H_{n\Lambda}$ and $H_{n\beta}$ with $H_{\Lambda\delta j}$ and $H_{\beta\delta k}$. Then for any $\eta > 0$

$$\text{(A26)}\qquad \lim_{n\to\infty} P\left[\sup_{u\in[u_0,u_1]} |E_n(u) - E_n^*(u)| > \eta\right]$$

$$\leq P\left[\max_{\substack{j\leq I_{\Lambda\delta}\\ k\leq I_{\beta\delta}}} \sup_{u\in[u_0,u_1]} |E_{njk}(u) - E_n^*(u)| > \eta\right] + \epsilon + O(\delta)$$

as $\delta \to 0$. Now use (A25) and the construction used to prove the Arzela theorem (see, e.g., Kolmogorov and Fomin (1970, p. 103)) to show that there are constants $\tilde{c}$ and $\tilde{d}$ such that $I_{\Lambda\delta} I_{\beta\delta} \leq \tilde{c}\delta^{\tilde{d}}$ for all sufficiently small δ. In addition, the functions $H_{\Lambda\delta i}$ are bounded. It follows that the class of functions of (y, x) indexed by (u, i, j, m) with typical element

$$1[\Lambda(y) - \beta' x + m^{-1/2} H_{\Lambda\delta i}(y) - m^{-1/2} H'_{\beta\delta j}\tilde{x} \leq u]$$

$$\cdot 1[-u \leq \beta' x + m^{-1/2} H'_{\beta\delta j}\tilde{x} \leq \Lambda(y_1) + m^{-1/2} H_{\Lambda\delta i}(y_1) - u]$$

is Euclidean. Therefore, by Lemma 2.16 of Pakes and Pollard (1989), the right-hand side of (A26) can be made arbitrarily small by making n sufficiently large and ϵ and δ sufficiently small. *Q.E.D.*

PROOF OF THEOREM 2b: By the mean value theorem of differential calculus

$$n^{1/2}[F_n(u) - F(u)]$$

$$= n^{1/2}[A_n(u)/B_n(u) - A(u)/B(u)]$$

$$\text{(A27)}\qquad = n^{1/2}[A_n(u) - A(u)]/B(u) - n^{1/2}[B_n(u) - B(u)]F(u)/B(u)$$

$$+ O\{n^{1/2}[A_n(u) - A(u)][B_n(u) - B(u)]\} + O\{n^{1/2}[B_n(u) - B(u)]^2\}.$$

1. Analysis of $n^{1/2}(B_n - B)$: By Lemma 9,

$$\text{(A28)}\qquad n^{1/2}[B_n(u) - B(u)] = e_n^*(u) + n^{1/2}\{\tilde{E}1[-u < b_n' x \leq \Lambda_n(y_1) - u]$$

$$- P[-u < Z \leq \Lambda(y_1) - u]\} + o_p(1)$$

uniformly over $u \in [u_0, u_1]$. Some algebra and a Taylor series expansion together with Lemma 7 show that

$$n^{1/2}\{\tilde{E}1[-u < b_n' x \leq \Lambda_n(y_1) - u] - P[-u < Z \leq \Lambda(y_1) - u]\}$$

$$= n^{1/2}[\Lambda_n(y_1) - \Lambda(y_1)]p[\Lambda(y_1) - u]$$

$$- n^{1/2}(\tilde{b}_n - \tilde{\beta})' \int \tilde{x}\{p[\Lambda(y_1) - u|\tilde{x}] - p(-u|\tilde{x})\} d\tilde{P}(\tilde{x}) + o_p(1)$$

$$\text{(A29)}\qquad = n^{-1/2}\sum_{i=1}^{n} J_{y_1}(Y_i, X_i)p[\Lambda(y_1) - u] - n^{-1/2}\sum_{i=1}^{n} \Omega(Y_i, X_i)'$$

$$\cdot \int \tilde{x}\{p[\Lambda(y_1) - u|\tilde{x}] - p(-u|\tilde{x})\} d\tilde{P}(\tilde{x}) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$. Substitution of (A29) into (A28) yields

$$\text{(A30)} \qquad n^{1/2}[B_n(u) - B(u)]$$

$$= e_n^*(u) + n^{-1/2} \sum_{i=1}^{n} J_{y_1}(Y_i, X_i) p[\Lambda(y_1) - u]$$

$$- n^{-1/2} \sum_{i=1}^{n} \Omega(Y_i, X_i)' \int \tilde{x}\{p[\Lambda(y_1) - u|\tilde{x}] - p(-u|\tilde{x})\} d\tilde{P}(\tilde{x}) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$. (A30) implies that $n^{1/2}[B_n(u) - B(u)] = O_p(1)$ uniformly over $u \in [u_0, u_1]$.

2. Analysis of $n^{1/2}[A_n(u) - A(u)]$: By Lemma 9

$$\text{(A31)} \qquad n^{1/2}[A_n(u) - A(u)] = E_n^*(u) + \mu_n(u) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$, where

$$\mu_n(u) = n^{1/2}\{\tilde{E}1[\Lambda_n(y) - b_n'x \le u]1[-u < b_n'x \le \Lambda_n(y_1) - u]$$

$$- E1[\Lambda(Y) - Z \le u]1[-u < Z \le \Lambda(y_1) - u]\}$$

$$\text{(A32)} \qquad = n^{1/2} \int P[\Lambda_n(Y) \le b_n'x + u|X = x]1[-u < b_n'x \le \Lambda_n(y_1) - u]\, dP(x)$$

$$- n^{1/2} F(u) P[-u < Z \le \Lambda(y_1) - u] + o_p(1)$$

uniformly over $u \in [u_0, u_1]$, where $P[\Lambda_n(Y) \le b_n'x + u|X = x]$ refers only to the distribution of Y and not to the distributions of b_n and Λ_n. To evaluate $P[\Lambda_n(Y) \le b_n'x + u|X = x] \equiv Q_n^*(x, u)$, define $y_n^* \equiv \inf\{y\colon\ \Lambda_n(y) > b_n'x + u,\ \ y_0 \le y \le y_1\} \equiv y_n^*(x, u)$. Also, define $y_\infty^* \equiv \Lambda^{-1}(\beta'x + u) \equiv y_\infty^*(x, u)$. Then

$$Q_n^*(x, u) \ge P(Y \le y_n^* | X = x)$$

$$= F[\Lambda(y_n^*) - \beta'x]$$

$$= F(u) + f(\tilde{u}_n)[\Lambda(y_n^*) - \lambda(y_\infty^*)],$$

where $\tilde{u}_n$ is between u and $\Lambda(y_n^*) - \beta'x$. In addition,

$$\Lambda(y_n^*) = \Lambda(y_\infty^*) + \lambda(\tilde{y}_n)(y_n^* - y_\infty^*),$$

where $\tilde{y}_n$ is between y_∞^* and y_n^*. Therefore

$$Q_n^*(x, u) \ge F(u) + f(\tilde{u}_n)\Lambda(\tilde{y}_n)(y_n^* - y_\infty^*).$$

Now since $\Lambda_n(\cdot)$ is a continuous function,

$$\Lambda_n(y_n^*) = \beta'x + u + (\tilde{b}_n - \tilde{\beta})'\tilde{x} = \Lambda(y_\infty^*) + (\tilde{b}_n - \tilde{\beta})'\tilde{x}.$$

Also, by the mean value theorem

$$\Lambda_n(y_n^*) = \Lambda_n(y_\infty^*) + \lambda_n(\bar{y}_n)(y_n^* - y_\infty^*),$$

where $\lambda_n(y) \equiv d\Lambda_n(y)/dy$ and $\bar{y}_n$ is between y_∞^* and y_n^*. Therefore,

$$y_n^* - y_\infty^* = -\lambda_n(\bar{y}_n)^{-1}\{[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)] - (\tilde{b} - \tilde{\beta})'\tilde{x}\}$$

and

$$Q_n^*(x,u) \geq F(u) - f(\tilde{u}_n)\lambda(\tilde{y}_n)\lambda_n(\bar{y}_n)^{-1}\{[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)] - (\tilde{b}_n - \tilde{\beta})'\tilde{x}\}.$$

Now $y_n^* \to y_\infty^*$ uniformly over $y_\infty^* \in [y_0, y_1]$ by uniform convergence of Λ_n to Λ. Therefore, $\tilde{y}_n \to y_\infty^*$ and $\bar{y}_n \to y_\infty^*$ uniformly. By arguments similar to those used in proving Theorem 1a, $\lambda_n(y) \to \lambda(y)$ uniformly over $y \in [y_0, y_1]$. Therefore, since $(\tilde{b}_n - \tilde{\beta}) = O_p(n^{-1/2})$ and $(\Lambda_n - \Lambda) = O_p(n^{-1/2})$ uniformly over $y \in [y_0, y_1]$,

$$Q_n^*(x,u) \geq F(u) - f(u)\{[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)] - (\tilde{b}_n - \tilde{\beta})'\tilde{x}\} + o_p(n^{-1/2};\tilde{x}),$$

uniformly over $u \in [u_0, u_1]$ and $y_\infty^* \in [y_0, y_1]$, where $o_p(n^{-1/2};\tilde{x})$ denotes a term whose size is $o_p(n^{-1/2}) + o_p(n^{-1/2})\|\tilde{x}\|$. Identical arguments with $y_n^* = \sup\{y\colon \Lambda_n(y) \leq b_n'x + u,\ y_0 \leq y \leq y_1\}$ yield

$$Q_n^*(x,u) \leq F(u) - f(u)\{[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)] - (\tilde{b}_n - \tilde{\beta})'\tilde{x}\} + o_p(n^{-1/2};\tilde{x})$$

uniformly over $u \in [u_0, u_1]$ and $y_\infty^* \in [y_0, y_1]$. Therefore

$$Q_n^*(x,u) = F(u) - f(u)\{[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)] - (\tilde{b}_n - \tilde{\beta})'\tilde{x}\} + o_p(n^{-1/2};\tilde{x})$$

uniformly over $u \in [u_0, u_1]$ and $y_\infty^* \in [y_0, y_1]$. Substituting these results into (A32) and making use of $(\Lambda_n - \Lambda) = O_p(n^{-1/2})$ and $(\tilde{b}_n - \tilde{\beta}) = O_p(n^{-1/2})$ yields

$$\mu_n(u) = F(u)\{n^{1/2}[B_n(u) - B(u)] - e_n^*(u)\} - f(u)\int\left\{\int_{-u}^{\Lambda(y_1)-u} n^{1/2}[\Lambda_n(y_\infty^*) - \Lambda(y_\infty^*)]p(z|\tilde{x})\,dz\right\}d\tilde{P}(\tilde{x}) + f(u)C_F'n^{1/2}(\tilde{b}_n - \tilde{\beta}) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$. A change of variables in the integral yields

$$\mu_n(u) = F(u)\{n^{1/2}[B_n(u) - B(u)] - e_n(u)\} - f(u)\int_{y_0}^{y_1} n^{1/2}[\Lambda_n(y) - \Lambda(y)] \cdot \lambda(y)p[\Lambda(y) - u]\,dy + f(u)C_F'n^{1/2}(\tilde{b}_n - \tilde{\beta}) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$. Substitution of this result into (A31) combined with Lemma 7 yields

$$\text{(A33)} \qquad n^{1/2}[\Lambda_n(u) - \Lambda(u)] = E_n^{*'}(u) + F(u)\{n^{1/2}[B_n(u) - B(u)] - e_n^*(u)\} - f(u)n^{-1/2}\sum_{\iota=1}^{n}\int_{y_0}^{y_1} J_y(Y_\iota, X_\iota)p[\Lambda(y) - u]\lambda(y)\,dy + f(u)C_F'n^{-1/2}\sum_{i=1}^{n}\Omega(Y_\iota, X_i) + o_p(1)$$

uniformly over $u \in [u_0, u_1]$.

3. Combining the results of (1) and (2): Substitution of (A33) and (A30) into (A27) yields

$$n^{1/2}[F_n(u)-F(u)]=n^{-1/2}\sum_{i=1}^{n}L_u(Y_i,X_i)+o_p(1) \tag{A34}$$

uniformly over $u\in[u_0,u_1]$. The functions $L_u(y,x)$ form a Euclidean class, so $n^{1/2}[F_n(u)-F(u)]\Rightarrow H_F$ follows from Theorem 5.3 of Pollard (1984). *Q.E.D.*

The following lemma is used in the proof of Theorem 3.

LEMMA 10: *Define u_γ and $u_{n\gamma}$ as in the text. If $u_\gamma\in[u_0+\epsilon,u_1-\epsilon]$ for some $\epsilon>0$ and $f(u_\gamma)>0$, then (a) $u_{n\gamma}\to^p u_\gamma$, and (b) $n^{1/2}(u_{n\gamma}-u_\gamma)=-n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]/f(u_\gamma)+o_p(1)$.*

PROOF: To prove part (a), define $\delta\equiv\min[F(u_\gamma)-F(u_\gamma-\epsilon/2),F(u_\gamma+\epsilon/2)-F(u_\gamma)]=\min[\gamma-F(u_\gamma-\epsilon/2),F(u_\gamma+\epsilon/2)-\gamma]>0$. If $u_0\le u<u_\gamma-\epsilon/2$, $F(u)\le\gamma-\delta$, and it follows from Theorem 2a that with probability approaching 1 as $n\to\infty$, $F_n(u)<\gamma-\delta/2$. If $u_1\ge u>u_\gamma+\epsilon/2$, $F(u)\ge\gamma+\delta$, and $F_n(u)>\gamma+\delta/2$ with probability approaching 1 as $n\to\infty$. These results imply that with probability approaching 1 as $n\to\infty$, $u_\gamma-\epsilon/2\le u_{n\gamma}\le u_\gamma+\epsilon/2$. Since ϵ is arbitrary, part (a) of the lemma is proved.

To prove part (b), note that since the class of functions $L_u(y,x)$ is Euclidean, it follows from (A34) and Lemma 2.16 of Pakes and Pollard (1989) that

$$n^{1/2}[F_n(u_{n\gamma})-F(u_{n\gamma})]=n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]+o_p(1).$$

Therefore,

$$\begin{aligned}n^{1/2}[F_n(u_{n\gamma})-F(u_\gamma)]&=n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]+n^{1/2}[F(u_{n\gamma})-F(u_\gamma)]+o_p(1)\\&=n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]+f(\tilde{u}_n)n^{1/2}(u_{n\gamma}-u_\gamma)+o_p(1),\end{aligned}$$

where $\tilde{u}_n$ is between u_γ and $u_{n\gamma}$. Since $F(u_\gamma)=\gamma$, $F_n(u_{n\gamma})\ge\gamma$ with probability approaching 1 as $n\to\infty$, $f(\tilde{u}_n)\to^p f(u_\gamma)$ and $n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]=O_p(1)$,

$$n^{1/2}(u_{n\gamma}-u_\gamma)\ge-n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]/f(u_\gamma)+o_p(1). \tag{A35}$$

By a similar argument

$$\begin{aligned}&n^{1/2}[F_n(u_{n\gamma}-1/n)-F(u_\gamma)]\\&\quad=n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]+n^{1/2}[F(u_{n\gamma}-1/n)-F(u_\gamma)]+o_p(1)\\&\quad=n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]+f(\bar{u}_n)n^{1/2}(u_{n\gamma}-u_\gamma)+o_p(1),\end{aligned}$$

where $\bar{u}_n$ is between u_γ and $u_{n\gamma}-1/n$. Since $F_n(u_{n\gamma}-1/n)<\gamma$ with probability approaching 1 as $n\to\infty$ and $f(\bar{u}_n)\to^p F(u_\gamma)$,

$$n^{1/2}(u_{n\gamma}-u_\gamma)<-n^{1/2}[F_n(u_\gamma)-F(u_\gamma)]/f(u_\gamma)+o_p(1). \tag{A36}$$

Part (b) follows by combining (A35) and (A36). *Q.E.D.*

PROOF OF THEOREM 3a: Define $S(\epsilon) \equiv [y_0 + \epsilon, y_1 - \epsilon]$. Let δ be the lesser of $\inf_{x:\, y(x) \in S(\epsilon)}\{\Lambda[y_\gamma(x) + \epsilon/2] - \beta' x - u_\gamma\}$ and $\inf_{x:\, y(x) \in S(\epsilon)}\{\beta' x + u_\gamma - \Lambda[y_\gamma(x) - \epsilon/2]\}$. Note that $\delta > 0$ since Λ is strictly increasing. If $y_0 \le y < y_\gamma(x) - \epsilon/2$ for some $x \in S(\epsilon)$, $\Lambda(y) \le \beta' x + u_\gamma - \delta$, and by Theorem 2a and Lemma 10 $\Lambda_n(y) < b_n' x + u_{n\gamma} - \delta/2$ with probability approaching 1 as $n \to \infty$. If $y_1 \ge y > y_\gamma(x) + \epsilon/2$ for some $x \in S(\epsilon)$, $\Lambda(y) \ge \beta' x + u_\gamma + \delta$, and by Theorem 2a and Lemma 10 $\Lambda_n(y) > b_n' x + u_{n\gamma} + \delta/2$ with probability approaching 1 as $n \to \infty$. Since $\Lambda_n \to \Lambda$ uniformly, it follows that $y_\gamma(x) - \epsilon/2 \le y_{n\gamma}(x) \le y_\gamma(x) + \epsilon/2$ with probability approaching 1 as $n \to \infty$ uniformly over x's such that $y \in S(\epsilon)$. Since ϵ is arbitrary, the theorem is proved. *Q.E.D.*

THEOREM 3b: *By Theorem* 3*a*, $y_{n\gamma}(x) \in [y_0, y_1]$ *uniformly over* $x \in S_{\epsilon\gamma}$ *with probability approaching* 1 *as* $n \to \infty$. *Therefore, since* $\Lambda_n \to^p \Lambda$ *uniformly*, $\Lambda_n[y_{n\gamma}(x)] = b_n' x + u_{n\gamma} + o_p(1)$ *and*

$$n^{1/2}\{\Lambda_n[y_{n\gamma}(x)] - \Lambda[y_{n\gamma}(x)]\} = n^{1/2}(\tilde{b}_n - \tilde{\beta})'\tilde{x} + n^{1/2}(u_{n\gamma} - u_\gamma)$$

$$- n^{1/2}\{\Lambda[y_{n\gamma}(x)] - \Lambda[y_\gamma(x)]\} + o_p(1)$$

uniformly over $x \in S_{\epsilon\gamma}$. *By Theorem* 3*a and Lemma* 2.16 *of Pakes and Pollard* (1989), $n^{1/2}\{\lambda_n[y_{n\gamma}(x)] - \Lambda[y_{n\gamma}(x)]\} = n^{1/2}\{\Lambda_n[y_\gamma(x)] - \Lambda[y_\gamma(x)]\} + o_p(1)$ *uniformly over* $x \in S_{\epsilon\gamma}$, *so*

$$n^{1/2}\{\Lambda_n[y_\gamma(x)] - \Lambda[y_\gamma(x)]\} = n^{1/2}(\tilde{b}_n - \tilde{\beta})'\tilde{x} + n^{1/2}(u_{n\gamma} - u_\gamma)$$

$$- n^{1/2}\{\Lambda[y_{n\gamma}(x)] - \Lambda[y_\gamma(x)]\} + o_p(1)$$

uniformly over $x \in S_{\epsilon\gamma}$. *A Taylor series expansion of* $\Lambda[y_{n\gamma}(x)]$ *followed by some algebra yields*

$$\begin{aligned} n^{1/2}[y_{n\gamma}(x) - y_\gamma(x)] &= \lambda[y_\gamma(x)]^{-1}\Big[n^{1/2}(\tilde{b}_n - \tilde{\beta})'\tilde{x} + n^{1/2}(u_{n\gamma} - u_\gamma) \\ &\qquad - n^{1/2}\{\Lambda_n[y_\gamma(x)] - \Lambda[y_\gamma(x)]\}\Big] + o_p(1) \\ &= \lambda[y_\gamma(x)]^{-1} n^{-1/2}\sum_{i=1}^{n}\Big\{\Omega(Y_i, X_i)'\tilde{x} - L_{u_\gamma}(Y_i, X_i) f(u_\gamma)^{-1} \\ &\qquad - J_{y_\gamma(x)}(Y_i, X_i)\Big\} + o_p(1) \end{aligned} \tag{A37}$$

uniformly over $x \in S_{\epsilon\gamma}$, *where Assumption* 6, *Lemma* 7, (*A*35) *and Lemma* 10*b have been used to obtain the last line. Convergence in distribution of* $n^{1/2}[y_{n\gamma}(x) - y_\gamma(x)]$ *follows from* (*A*37), *the continuous mapping theorem, and joint convergence in distribution of* $n^{1/2}(\tilde{b}_n - \tilde{\beta})$, $n^{1/2}(u_{n\gamma} - u_\gamma)$, *and* $n^{1/2}(\Lambda_n - \Lambda)$. *The covariance function follows from* (*A*37). *Q.E.D.*

APPENDIX B: ESTIMATION OF COVARIANCE FUNCTIONS

a. Estimation of $E[H_\Lambda(y)H_\Lambda(y')]$

By Theorem 1, $E[H_\Lambda(y)H_\Lambda(y')] = E_{YX}[J_y(Y, X)J_{y'}(Y, X)]$. To obtain a consistent estimator of $E_{YX}[J_y(Y, X)J_{y'}(Y, X)]$, define $p_n'(z) = dp_n(z)/dz$, $g_{nz}(y, z) = G_{nz}(y|z)p_n(z)$, $\lambda_n(v) = d\Lambda_n(v)/dv$, $D_n(y, z) = \partial[w(z)/g_{nz}(y, z)]/\partial z$, and

$$C_{n\Lambda} = \int [w'(z)\tilde{x}p_n(z|\tilde{x})/p_n(z)]\, dz\, d\tilde{P}_n(\tilde{x}),$$

where $p_n(z|\tilde{x})$ is a kernel estimator of $p(z|x)$ and $\tilde{P}_n$ is the EDF of $\tilde{X}$. Let Ω_n be a consistent estimator of Ω. Ω_n depends on how β is estimated. Formulae for calculating Ω_n's corresponding to a variety of estimators of β are given in the references in Section 1 of this paper.

J_y can be estimated by replacing population quantities in the expression for J_y with the foregoing estimators. The result is

$$\begin{aligned}
&J_{ny}(Y,X)\\
&\quad\equiv -w(Z_n)\left[\frac{1(y_0<Y\leq y)-1(y<Y\leq y_0)}{g_{nz}(Y,Z_n)}+\frac{\Lambda_n(y)}{p_n(Z_n)}\right]\\
&\quad\quad+\int_{y_0}^{y}\lambda_n(v)\left[\frac{w(Z_n)p_n'(Z_n)}{g_{nz}(v,Z_n)p_n(Z_n)}+D_n(v,Z_n)\right][1(Y\leq v)-G_n(v|Z_n)]\,dv\\
&\quad\quad-\Lambda_n(y)C_{n\Lambda}'\Omega_n(Y,X).
\end{aligned}$$

Under the regularity conditions of Theorem 1, $J_{ny}(Y,X)=J_y(Y,X)+o_p(1)$. It can be shown that for $(y,y')\in[y_2,y_1]$, $E_{YX}[J_y(Y,X)J_{y'}(Y,X)]$ is estimated consistently by

$$n^{-1}\sum_{i=1}^{n}J_{ny}(Y_i,X_i)J_{ny'}(Y_i,X_i).$$

b. *Estimation of* $E[H_F(y)H_F(y')]$

By Theorem 2, $E[H_F(u)H_F(u')]=E_{YX}[L_u(Y,X)L_{u'}(Y,X)]$. To develop a consistent estimator of $E_{YX}[L_u(Y,X)L_{u'}(Y,X)]$, let P_n be the EDF of $Z_n\equiv b_n'X$, and $P_n(z|\tilde{x})$ be a uniformly consistent estimator of the CDF of Z conditional on $\tilde{X}=\tilde{x}$. $P_n(z|\tilde{x})$ may be obtained, among other ways, by integrating $p_n(z|\tilde{x})$. Let $f_n(u)$ be an estimator of $f(u)$ that is consistent uniformly over $u\in[u_0+\epsilon,u_1-\epsilon]$ for any $\epsilon>0$. One possible f_n is the kernel estimator

$$f_n(u)=(1/h)\int_{u_0}^{u_1}K\left(\frac{u-v}{h}\right)dF_n(v),$$

where K is a continuous, symmetrical probability density function on $[-1,1]$ and h is a bandwidth. Define $\pi_n(u)=P_n[\Lambda_n(y_1)-u]-P_n[\Lambda_n(y_2)-u]$,

$$C_{nF}(u)=\int\tilde{x}P_n[\Lambda_n(y_2)-u<Z_n\leq\Lambda_n(y_1)-u|\tilde{x}]\,d\tilde{P}_n(\tilde{x}),$$

and

$$\begin{aligned}
&L_{nu}(Y,Z)\\
&\quad=\pi_n(u)^{-1}\Big\{\{1[\Lambda_n(Y)\leq Z_n+u]-F_n(u)\}1[\Lambda_n(y_2)\\
&\quad\quad-u<Z_n\leq\Lambda_n(y_1)-u]\\
&\quad\quad-f_n(u)\int_{y_2}^{y_1}J_{ny}(Y,X)p_n[\Lambda_n(y)-u]\lambda_n(y)\,dy\\
&\quad\quad+f_n(u)C_{nF}(u)'\Omega_n(Y,X)\Big\}.
\end{aligned}$$

Under the assumptions of Theorem 2, $C_{nF}(u) = C_F(u) + o_p(1)$ and $L_{nu}(Y, X) = L_u(Y, X) + o_p(1)$ uniformly over $u \in [u_0 + \epsilon, u_1 - \epsilon]$. It can be shown that for $(u, u') \in (u_0, u_1)$, $E_{YX}[L_u(Y, X) L_{u'}(Y, X)]$ is estimated consistently by

$$n^{-1} \sum_{i=1}^{n} L_{nu}(Y_i, X_i) L_{u'}(Y_i, X_i).$$

c. Estimation of $E[H_\gamma(x) H_\gamma(x')]$

By Theorem 3, $E[H_\gamma(x) H_\gamma(x')]$ is the same as the covariance function of

$$M_x(Y, X) \equiv \lambda[y_\gamma(x)]^{-1} \Big\{ \Omega(Y, X)' \tilde{x} - L_{u_\gamma}(Y, X) f(u_\gamma)^{-1} - J_{y_\gamma(x)}(Y, X) \Big\}.$$

Define

$$M_{nx}(Y, X) \equiv \lambda_n[y_{n\gamma}(x)]^{-1} \Big\{ \Omega_n(Y, X)' \tilde{x} - L_{nu_\gamma}(Y, X) f_n(u_{n\gamma})^{-1} - J_{n y_{n\gamma}(x)}(Y, X) \Big\}.$$

For $(x, x') \in S_{\epsilon\gamma}$, $E[H_\gamma(x) H_\gamma(x')]$ is estimated consistently by

$$n^{-1} \sum_{i=1}^{n} M_{nx}(Y_i, X_i) M_{nx'}(Y_i, X_i).$$

REFERENCES

AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

BICKEL, P. J., AND K. A. DOKSUM (1981): "An Analysis of Transformations Revisited," *Journal of the American Statistical Association*, 76, 296–311.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: John Wiley & Sons.

BOX, G. E. P., AND D. R. COX (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Series B, 26, 211–243.

BREIMAN, L., AND J. H. FRIEDMAN (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580–598.

COX, D. R. (1972): "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, Series B, 34, 187–220.

ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazards Model," *Review of Economic Studies*, 49, 402–411.

HAN, A. K. (1987a): "Non-Parametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

——— (1987b): "A Non-Parametric Analysis of Transformations," *Journal of Econometrics*, 35, 191–209.

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.

HASTIE, T. J., AND R. J. TIBSHIRANI (1990): *Generalized Additive Models*. New York: Chapman and Hall.

HECKMAN, J. J., AND B. SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

ICHIMURA, H. (1993): "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics*, 58, 71–120.

JOHN, J. A., AND N. R. DRAPER (1980): "An Alternative Family of Transformations," *Applied Statistics*, 29, 190–197.

JOHNSON, N. L. (1949): "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149–176.

KOLMOGOROV, A. N., AND S. V. FOMIN (1970): *Introductory Real Analysis*. New York: Dover Publications.

MACKINNON, J. G., AND L. MAGEE (1990): "Transforming the Dependent Variable in Regression Models," *International Economic Review*, 31, 315–339.

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.

——— (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," *Scandanavian Journal of Statistics*, 19, 25–43.

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

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): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 474–523.

RIDDER, G. (1990): "The Non-Parametric Identification of Generalized Accelerated Failure-Time Models," *Review of Economic Studies*, 57, 167–182.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–1954.

——— (1991): "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica*, 59, 755–786.

SHERMAN, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.