The Annals of Statistics
1997, Vol. 25, No. 6, 2682–2717

NONPARAMETRIC $n^{-1/2}$-CONSISTENT ESTIMATION FOR THE GENERAL TRANFORMATION MODELS

By Jianming Ye[1] and Naihua Duan[2]

University of Chicago and RAND Corporation

We propose simple estimators for the transformation function Λ and the distribution function F of the error for the model

$$\Lambda(Y) = \alpha + \mathbf{X}\boldsymbol{\beta} + \varepsilon.$$

It is proved that these estimators are consistent and can achieve the unusual $n^{-1/2}$ rate of convergence on any finite interval under some regularity conditions. We show that our estimators are more attractive than another class of estimators proposed by Horowitz. Interesting decompositions of the estimators are obtained. The estimator of F is independent of the unknown transformation function Λ, and the variance of the estimator for Λ depends on Λ only through the density function of X. Through simulations, we find that the procedure is not sensitive to the choice of bandwidth, and the computation load is very modest. In almost all cases simulated, our procedure works substantially better than median nonparametric regression.

1. Introduction. Consider the model

$$\Lambda(Y) = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{1}$$

where $\Lambda(y)$ is a strictly increasing function of y and ε is random noise with an unknown distribution F, and is independent of the random vector $\mathbf{X}$. Following Efron (1982), we refer to model (1) as the *general transformation model*. For different forms of Λ and F, this model generates many different parametric families of models. For example, if Λ takes the form of a power function and F a normal distribution, (1) reduces to the familiar Box–Cox transformation models [see Box and Cox (1964, 1982), Bickel and Doksum (1981) and Hinkley and Runger (1984)]. Many models related to this family are given in Caroll and Ruppert (1988). Another important subfamily is the parametric proportional hazards models where F takes an extreme value distribution [see Doksum (1987)] and their extension to proportional hazards models with random effects [Heckman and Singer (1984) and Murphy (1994, 1995)]. In this paper, we will primarily study the estimation of Λ and F assuming that $\boldsymbol{\beta}$ is a known quantity, and give some simulation results for

Received February 1994; revised March 1997.
[1]Supported in part by the University of Chicago Graduate School of Business Research Fund.
[2]Supported in part by the RAND Corporation research funds.
AMS 1991 *subject classifications*. General transformation models, median nonparametric regression, shifted median estimator, uniform convergence, mean integrated square error, prediction interval.
Key words and phrases. Primary 62G07; secondary 62G02.

the case of unknown $\boldsymbol{\beta}$. To make the model identifiable, we will assume that $F(0) = 0.5$ and $\mathbf{Z} = X\boldsymbol{\beta}$ has nonzero variance.

Several important contributions were made recently in nonparametric estimation of Λ and F that achieves $n^{-1/2}$ rate of convergence, a result relatively unusual in estimating nonparametric components. Wang and Ruppert (1995) studied the transform-both-sides model, a model closely related to (1),

$$\Lambda(Y) = \Lambda(u(\mathbf{X}, \boldsymbol{\beta})) + \varepsilon.$$

They showed that a root-n estimation for Λ is possible by first estimating its derivative Λ'. A recent work by Horowitz (1996) proposed a similar type of estimator for both Λ and F in (1) and established the result of the $n^{-1/2}$ consistency and asymptotic normality for these estimators. Unfortunately, Horowitz's estimators suffer several serious drawbacks and they perform poorly in practice. The goal of this paper is to provide alternative $n^{-1/2}$-consistent estimators for Λ and F to overcome the difficulties faced by Horowitz's estimators.

To estimate Λ, Efron (1982), Wang and Ruppert (1995) and Horowitz (1996) study the partial first-order derivative of the conditional density of Y given X. Let $Z = \mathbf{X}\boldsymbol{\beta}$ in (1). Efron (1982) and Horowitz (1996) find that

$$\Lambda'(y) = -\frac{G_1'(y \mid z)}{G_2'(y \mid z)}$$

holds for every (z, y), where $G_1'(y \mid z)$ and $G_2'(y \mid z)$ are the first-order partial derivatives of $G(Y \mid Z)$, the distribution function of Y conditioned on Z, with respect to Y and Z, respectively. Based on this relationship, Horowitz (1996) estimates Λ by

$$\hat{\Lambda}(y) = \int_{y_0}^{y} \int_{S_z} w(z) \frac{\hat{G}_1'(v \mid z)}{\hat{G}_2'(v \mid z)}\, dz\, dv \tag{2}$$

by using the estimates of partial derivatives, with a weight function $w(z)$ that has support on S_z. Horowitz established the result of $n^{-1/2}$ rate of convergence for $\hat{\Lambda}$ and $\hat{F}$. However, the estimator (2) has several shortcomings. One major problem is in its use of derivatives in the estimation, especially in the denominator. It is well known that the derivatives are difficult to estimate, especially in small-sample situations. The estimator may be highly unstable if $\hat{G}_2'(v \mid z)$ is near 0. Even remedies such as thresholding are inadequate to solve the problem. The second drawback for this estimator is its integral form (or CUSUM form). Based on such a construction, an estimation error in $G_1'(y \mid z)/G_2'(y \mid z)$ for y^* would persist for all $y \geq y^*$. Therefore, the result of $n^{-1/2}$ consistency and asymptotic normality requires the existence of very high order derivatives (up to the ninth order) of the unknown quantities. The requirement for high-order derivatives is mostly theoretical, but it implies that the underlying functions have to be very smooth. Due to these difficulties, the estimator performs poorly in practice.

The method we propose in this paper is related to that of Doksum (1987), who utilizes the additive structure (or ANOVA structure) of the general transformation models, that is,

$$\Lambda(y) = z + F^{-1}(G(y \mid z)) \tag{3}$$

and estimates Λ by replacing $G(y \mid z)$ by its estimate $\hat{G}(y \mid z)$ and averaging over different z's, assuming known F. In this paper, we propose an estimator for F that achieves the $n^{-1/2}$ rate and then estimate $\Lambda(y)$ by replacing F in (3) by $\hat{F}$. Therefore, we focus on the estimation of F and its influences on the estimation of $\Lambda(y)$. The estimators are shown to overcome the difficulties with Horowitz's estimators and perform quite well in simulations.

We also carry out simulations for the more realistic case of unknown $\boldsymbol{\beta}$, and the improvement based on the new estimator is still quite substantial. The coefficient $\boldsymbol{\beta}$ can be obtained in two different ways. A straightforward procedure is to plug in an $n^{-1/2}$-consistent estimator for the direction of $\boldsymbol{\beta}$, such as a projection pursuit regression estimate [Friedman and Stuetzle (1981), Duan (1990) and Chaudhuri, Doksum and Samarov (1994)] or the average derivative estimate (ADE, Härdle and Stoker (1989)]. Due to the additive structure of our estimators, the error in estimating Λ and F should still be of the same order. A potentially better approach utilizes the estimate of Λ and F obtained here and obtains $\hat{\boldsymbol{\beta}}$ by a pseudo-MLE method as established in Wang and Ruppert (1996) for the transform-both-sides models.

The rest of the paper is arranged as follows. Our estimators are proposed in Section 2. Section 3 deals with the theoretical aspects of them. Simulation results and discussions on the implementation details of the algorithm are presented in Section 4. In Section 5, several alternative algorithms are given and discussed.

2. The basic algorithm. Assume that $z_1, z_2, \ldots, z_n$ are distributed with a density function $d(z)$ which is supported on the interval $[a, b]$. Without loss of generality, we assume that $a = 0$ and $b = 1$. Let $G(y \mid z)$ be the distribution function of y given z and $q_p(z)$ be its pth quantile. To motivate our estimators, note that

$$q_p(z) = \Lambda^{-1}\big(z + F^{-1}(p)\big),$$

which is another expression of (3). Using the initial condition that $F^{-1}(0.5) = 0$, we have

$$\Lambda\big(q_{0.5}(z)\big) = z,$$

which implies that, for $0 \le \Delta < 1$ and $z \in [0, 1 - \Delta]$,

$$\begin{aligned} F(-\Delta) &= F\big(\Lambda(q_{0.5}(z)) - (z + \Delta)\big) \\ &= P\big(\varepsilon \le \Lambda(q_{0.5}(z)) - (z + \Delta)\big) \\ &= P\big((z + \Delta) + \varepsilon \le \Lambda(q_{0.5}(z))\big) \\ &= G\big(q_{0.5}(z) \mid z + \Delta\big). \end{aligned} \tag{4}$$

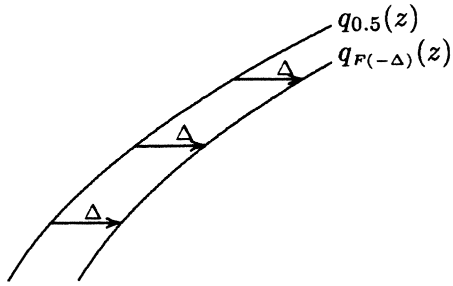

FIG. 1. *Illustration of* (4).

As Figure 1 illustrates, the error distribution is determined by shifting the conditional medians of y given z. Since the same quantity $F(-\Delta)$ can be determined from (4) using different z's, we can average the right-hand side of (4) across z to combine the information:

$$F(-\Delta) = \frac{1}{1-\Delta}\int_0^{1-\Delta} P(y \le q_{0.5}(z) \mid z + \Delta)\, dz. \tag{5}$$

Note that (4) is valid only for z between 0 and $1 - \Delta$. Similar expressions can be derived for $-1 < \Delta < 0$ and for other quantiles.

Let $\delta_n = [cn^r]$ be a sequence of positive integers, where r falls between 1/4 and 1/3 for reasons to be discussed in Section 3, and $[a]$ indicates the largest integer less than or equal to a. Let $z_{j,n}^0 = (j - 1/2)/\delta_n$ for $j = 1, 2, \dots, \delta_n$ be a grid of points in $[0, 1]$. Denote by $I_n(z)$ the interval $((z - 1/2)1/\delta_n, (z + 1/2)1/\delta_n)$. If the context is clear, we will suppress the subscript n for $z_{j,n}^0$ and $I_n(z)$, and use z_j^0 and $I(z)$ instead. By expression (5), the cumulative error distribution function $F(-\Delta)$ can be estimated by

$$\hat{F}(-\Delta) = \frac{1}{[\delta_n(1-\Delta)]}\sum_{j=1}^{[\delta_n(1-\Delta)]} \frac{\#\left\{y_i \le \hat{q}_{0.5}\left(z_j^0\right), z_i \in I\left(z_j^0 + \Delta\right)\right\}}{\#\left\{I\left(z_j^0 + \Delta\right)\right\}} \tag{6}$$

for $0 \le \Delta < 1$, where $\#\{S\}$ is the number of observations in the set S and $\hat{q}_{0.5}(z_j^0)$ is the sample median of y in the interval $I(z_j^0)$. For $-1 < \Delta < 0$, a similar estimate for $F(-\Delta)$ is given by

$$\hat{F}(-\Delta) = \frac{1}{\delta_n(1-\Delta)}\sum_{j=[\delta_n\Delta]}^{\delta_n} \frac{\#\left\{y_i \le \hat{q}_{0.5}\left(z_j^0\right), z_i \in I\left(z_j^0 - \Delta\right)\right\}}{\#\left\{I\left(z_j^0 - \Delta\right)\right\}}. \tag{6$'$}$$

Since $\hat{F}(0) = 0.5$, the condition that $\hat{F}^{-1}(0.5) = 0$ is automatically satisfied.

Another interesting property of this estimator is that it is independent of the transformation Λ. This is best seen by converting the raw data into ranks. Since the quantiles and percentiles depend only on ranks and the ranks are invariant under monotone transformations, $\hat{F}$ is independent of Λ.

The medians $q_{0.5}(z_j^0)$ and $\hat{q}_{0.5}(z_j^0)$ are used in (4), (5), (6) and (6′). Any other quantiles $q_p(z_j^0)$ and $\hat{q}_p(z_j^0)$, $p \in (0,1)$ can also be used. From Theorem 1 in Section 3, one will see that the median is usually a reasonable choice. In our simulations, it is used for all cases. Different choices of $\hat{q}_{0.5}(z_j^0)$, $j = 1, \dots, \delta_n$, may be used in (6). One may apply a monotonic smoothing to $\hat{q}_{0.5}(z_j^0)$ to ensure that $\hat{q}_{0.5}(z_j^0)$ is monotonically increasing in j.

To estimate F consistently on an arbitrarily wide interval, we can "splice" the preceding estimate using different quantiles. (The shifted median estimator proposed previously can be used only on the range $-1 < \Delta < 1$.) More specifically, we have the following algorithm:

(A1) Choose a $0 < \eta < 1$.

(A2) Estimate $F(-\Delta)$, $\Delta \in [-\eta, \eta]$, using medians by the preceding algorithm.

(A3) Estimate $F(-\Delta)$, $\Delta \in (k\eta, (k+1)\eta]$, $k = 1, 2, \dots,$ by

$$(7) \quad \hat{F}(-\Delta) = \frac{1}{[\delta_n(1-\Delta)]} \sum_{j=1}^{[\delta_n(1-\Delta)]} \frac{\#\left\{y_i \le \hat{q}_{p_k}\left(z_j^0\right),\, z_i \in I\left(z_j^0 + \Delta\right)\right\}}{\#\left\{I\left(z_j^0 + \Delta\right)\right\}},$$

using the p_kth quantile, where $p_k = \hat{F}(-k\eta)$ is estimated from the previous step.

(A4) Similarly, estimate $F(-\Delta)$, $\Delta \in ((-k-1)\eta, -k\eta]$, $k = 1, 2, \dots,$ by

$$\hat{F}(-\Delta) = \frac{1}{\delta_n(1-\Delta)} \sum_{j=[\delta_n \Delta]}^{\delta_n} \frac{\#\left\{y_i \le \hat{q}_{p_k}\left(z_j^0\right),\, z_i \in I\left(z_j^0 - \Delta\right)\right\}}{\#\left\{I\left(z_j^0 - \Delta\right)\right\}},$$

using the p_kth quantile, where $p_k = \hat{F}(k\eta)$ is estimated from the previous step.

(A5) Repeat until $k = k_0$ is large enough so that $\hat{F}(-k_0\eta)$ and $\hat{F}(k_0\eta)$ are sufficiently close to 0 and 1, respectively.

The only requirement in (A1)–(A5) is that η should not be too close to either 0 or 1. If the distribution of Z is symmetric, a sensible choice for η would be median(Z)—at least half of the observations are available for estimating $F(\Delta)$ for Δ in the "basic range" $[-\eta, \eta]$.

To estimate $\Lambda(y)$, observe that (1) implies (3) for all z. This relationship is used in Doksum (1987) to define his estimator for Λ for known F. Averaging (3) over $z \in [0,1]$, we have

$$\Lambda(y) = \int_0^1 \left(z + F^{-1}(G(y \mid z))\right) dz.$$

Therefore, to estimate $\Lambda(y)$, we simply replace F^{-1} and $G(y \mid z)$ by their estimates. For a preselected $0 < \gamma < k_0\eta$, the estimator for $\Lambda(y)$ is then defined to be

$$\hat{\Lambda}(y) = \frac{1}{\Sigma w\left(\hat{G}\left(y \mid I\left(z_j^0\right)\right)\right)} \times \sum_{j=1}^{\delta_n} w\left(\hat{G}\left(y \mid I\left(z_j^0\right)\right)\right)\left[z_j^0 + \hat{F}^{-1}\left(\hat{G}\left(y \mid I\left(z_j^0\right)\right)\right)\right], \tag{8}$$

where

$$\hat{G}(y \mid I(z)) = \frac{\#\{y_i \le y,\, z_i \in I(z)\}}{\#\{I(z)\}}, \qquad w(p) = \begin{cases} 1, & \text{if } p \in \left(\hat{F}(-\gamma), \hat{F}(\gamma)\right), \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

The role of the weight function $w(p)$ here is to control the behavior of the estimator and to avoid the difficulty in estimating $F^{-1}(p)$ for very small or large p. If γ is chosen close to 0, then only the quantile information of F near $p = 0.5$ is used, and $\hat{\Lambda}$ would be close to the median nonparametric regression. The larger the γ is, the more quantile information is used. On the other hand, since $F^{-1}(p)$ cannot be estimated well for p near 0 or 1, increasing γ over a certain limit may only increase the noise. Thus, a reasonable selection of γ would be important. Based on our limited experience, however, other nonuniform weighting functions different from $w(p)$ given previously may be better. For example, letting $p_j = \hat{G}(y \mid I(z_j^0))$, the weighting function

$$w(p_j) = \begin{cases} \left(\#\left\{I\left(z_j^0\right)\right\}\right)p_j(1 - p_j), & \text{if } p_j \in \left(\hat{F}(-\gamma), \hat{F}(\gamma)\right), \\ 0, & \text{otherwise} \end{cases} \tag{9'}$$

would provide appropriate adjustment to the variance of $\hat{G}(y \mid z)$.

For the purpose of prediction, one can obtain the conditional median of y given z by inverting $\hat{\Lambda}$:

$$\hat{\Lambda}^{-1}(z) = \inf\left\{y \in [y_0, y_1],\, \hat{\Lambda}(y) \le z\right\},$$

where y_0 and y_1 are two preselected constants defining the range of the possible values of $\hat{\Lambda}^{-1}$.

REMARK. A key issue encountered in the estimation is the selection of the bandwidth, that is, the width of intervals for data grouping. From the theoretical results in Section 3, the number of intervals should be proportional to n^r, with r being in the interval (1/4, 1/3). Thus, the estimate is relatively robust for different choice of r. We will show this later in the simulation studies.

Instead of estimating $G(y \mid z)$ by $\hat{G}$ in (9), alternative estimates may be used, such as an isotonic smoothing of $\hat{G}_j(y \mid z_j^0)$, $j = 1, \dots, \delta_n$, using the number of observations $\#\{I(z_j^0)\}$ as the weights, so that $\hat{G}_j(y \mid z_j^0)$ is monotonically decreasing in j. In the simulations, the isotonization often improves the final estimates for Λ and F substantially. The estimates $\hat{F}$ and $\hat{\Lambda}$ can be also isotonized to guarantee their monotonicity.

A recursive procedure that reiterates the new procedure with the median estimates from the previous cycle may lead to further improvement of the final estimates. We will not pursue this fine tuning here.

3. Theoretical results. In this section, we present the theoretical results on the estimators of Λ and F proposed in Section 2. In order to make the result more transparent, we rewrite (6) in a slightly different form

$$\hat{F}(-\Delta) = \frac{1}{\delta_n(1-\Delta)} \sum_{j=1}^{[\delta_n(1-\Delta)]} \frac{\#\left\{y_i \le \hat{q}_{0.5}\left(z_j^0\right),\, z_i \in I\left(z_j^0 + \Delta\right)\right\}}{\#\left\{I\left(z_j^0 + \Delta\right)\right\}},$$

where we ignore the difference between $[\delta_n(1-\Delta)]$ and $\delta_n(1-\Delta)$ since it is of a negligible order $o(n^{-1/2})$. The following assumptions are needed to obtain the desired theoretical results.

Assumption 1. $\{z_i\}_{i=1}^{i=+\infty}$ is a sequence of independently identically distributed (iid) random variables with the density function $d(z)$ which is bounded away from 0 and $+\infty$ on $[0,1]$, that is, $0 < m_d < d(z) < M_d < +\infty$.

Assumption 2. (z_i, y_i), $i = 1, 2, \ldots,$ are iid realizations of model (1).

Assumption 3. $\Lambda(y)$ is twice differentiable, and

(a) $\lambda(y) = \Lambda'(y)$ satisfies the condition that $0 < m_\lambda < \lambda(y) < M_\lambda < +\infty$ on the interval $[\Lambda^{-1}(0) - K, \Lambda^{-1}(1) + K]$ for some positive constant K.

(b) $\Lambda''(y)$ is bounded from above on the same interval.

Assumption 4. F is twice differentiable, and

(a) $f(\varepsilon) = F'(\varepsilon)$ exists and is bounded away from 0 and $+\infty$, that is, $0 < m_f \le f(z) \le M_f < +\infty$ on the interval $(-\delta_1, \delta_2)$ for some constants $\delta_1, \delta_2 > 0$.

(b) $F''(z)$ is bounded from above.

The main role of the boundedness conditions in Assumptions 1, 3 and 4 is to guarantee that Λ and F do not have zero or infinite slope. An infinite slope in Λ or a zero slope in F might lead to difficulty or nonuniqueness in estimating the conditional quantile of $G(y \mid z)$. A zero slope in Λ or an infinite slope in F (a jump) would generate a point mass which would lead to nonseparability in quantile estimation.

The second-order differentiability requirement is needed for the existence of the Bahadur representation of the conditional quantiles.

ASSUMPTION 5. $1/4 < r < 1/3$, where r is the rate the bandwidth $1/\delta_n$ shrinks: $\delta_n = [cn^r]$.

This assumption is a major departure from most existing literature in nonparametric regression, including the class of estimators proposed by Horowitz (1996). It indicates that the bandwidth selection is not as important as it is in most other cases. The upper bound, $r = 1/3$, is to guarantee that the estimated quantiles will behave well enough so that the reminder term in the Bahadur representation would be negligible; and the lower bound, $r = 1/4$, is assumed so that the distribution of y in each subinterval approximates the conditional distribution $G(y \mid z)$ well enough. The reason for the insensitivity to the bandwidth is due to aggregation of the quantile curves. However, the estimators are not totally independent of the bandwidth. Obviously, if $r = 1$, there would be too few points in each interval such that $\hat{G}(y \mid z) = 0$ or 1 for all z values. If $r = 0$, the approximation is too crude and so the bias is too large.

This assumption is an important result itself. It indicates that the effect of bandwidth is only of secondary importance within a reasonable range. Therefore, for some constant c, if the number of intervals is between $3.16c$ and $4.64c$ for $n = 100$, or between $5.62c$ and $10c$ for $n = 1000$, the performance of the estimators is affected only to an order smaller than the main error term.

We will discuss the assumptions in more detail later in the sketch of the proofs. Under the preceding assumptions, we can establish the convergence properties of the estimators of Λ and F.

THEOREM 1. *For any given constant* $0 \leq \eta < 1$, *we have*

$$\operatorname*{plim}_{n\to\infty} \sup_{-\eta \leq \Delta \leq 0} |\hat{F}(\Delta) - F(\Delta)| = 0.$$

Moreover,

$$n^{1/2}\left[\hat{F}(\Delta) - F(\Delta)\right] \Rightarrow_{\mathscr{L}} H_1(\Delta)$$

for $-\eta \leq \Delta \leq 0$, *where* $\Rightarrow_{\mathscr{L}}$ *indicates process convergence*, $H_1(\Delta)$ *is a zero-mean tight Gaussian process with covariance function*

$$\begin{aligned}
R_1(s,t) &= \operatorname{Cov}(H_1(s), H_1(t)) \\
&= \frac{1}{(1+s)(1+t)} \\
&\quad \times \Bigg\{ F(t)(1-F(s)) \int_t^1 \frac{1}{d(z)}\,dz + \frac{f(s)f(t)}{f^2(0)} \frac{1}{4} \int_0^{1+t} \frac{1}{d(z)}\,dz \\
&\qquad - 1(s+t > -1)\Bigg[\frac{f(s)F(t)}{2f(0)} \int_t^{1+s} \frac{1}{d(z)}\,dz \\
&\qquad\qquad + \frac{f(t)F(s)}{2f(0)} \int_s^{1+t} \frac{1}{d(z)}\,dz \Bigg]\Bigg\},
\end{aligned} \tag{10}$$

where $-\eta \le t \le s \le 0$. *Furthermore*, $[\hat{F}(\Delta) - F(\Delta)]$ *has the representation*

$$\hat{F}(\Delta) - F(\Delta) = n^{-1/2}H_{11}^{*}(\Delta) + n^{-1/2}H_{12}^{*}(\Delta; 0.5) + o(n^{-1/2}), \tag{11}$$

where

$$H_{11}^{*}(\Delta) = \frac{n^{-1/2}}{(1+\Delta)} \sum_{z_i \in (-\Delta, 1)} \frac{1(\varepsilon_i \le \Delta) - F(\Delta)}{d(z_i)},$$

$$H_{12}^{*}(\Delta; p) = -\frac{n^{-1/2} f(\Delta)}{f(F^{-1}(p))(1+\Delta)} \sum_{z_i \in (0, (1+\Delta))} \frac{1(\varepsilon_i \le F^{-1}(p)) - p}{d(z_i)},$$

and $1(x < y)$ *is the indicator function for* $x < y$.

The representation (11) consists of two major additive parts: the first part is the estimation error when the medians are known exactly, which is comparable in size to the estimation error for the ordinary empirical distribution; the second part is the error induced from the estimation of the medians, which is the loss in accuracy we sustain for not knowing $\Lambda(y)$. A nice property of this estimator is that the distribution of $n^{1/2}[\hat{F}(\Delta) - F(\Delta)]$ is independent of the function Λ, as discussed in Section 2. We can also see this property from the asymptotic representation. It is worth noting that $R_1(0,0) = 0$, which is consistent with the initial condition that $\hat{F}(0) = 0.5$. Furthermore, the covariance function has a rather simple form and can be estimated consistently.

Since we use the shifted-median estimator, the denominator for the second part of (11) is $f(0)$. If we estimate F by shifting the pth quantile, it would be $f(F^{-1}(p))$. This suggests that the shifted-mode estimator, if the mode is known, may be better to use. Another inflator of the overall error in (11) is $1/(1+\Delta)$ for $\Delta \in [-\eta, 0]$. Thus, it is important to keep η away from 1 and use the recursive algorithm (A1)–(A5) if needed.

Results similar to Theorem 1 can be obtained for $\hat{F}(\Delta)$, $\Delta \in [0, \eta]$. For the recursive algorithm (A1)–(A5), we have the following corollary:

COROLLARY 1. *Let* $\hat{F}(\Delta)$, $\Delta \in [-(k+1)\eta, -k\eta)$, *be obtained from* (7) *for* $k = 1, 2, \dots, K$,

$$\operatorname*{plim}_{n\to\infty} \sup_{-(k+1)\eta \le \Delta < -k\eta} |\hat{F}(\Delta) - F(\Delta)| = 0.$$

Moreover,

$$n^{1/2}\big[\hat{F}(\Delta) - F(\Delta)\big] \Rightarrow_{\mathscr{L}} H_1(\Delta)$$

for $-(k+1)\eta < \Delta < -k\eta$, *where* $H_1(\Delta)$ *is a zero-mean tight Gaussian process. Furthermore,*

$$
\begin{aligned}
\hat{F}(\Delta) - F(\Delta) = n^{-1/2}\Bigg[& H_{11}^*(\Delta + k\eta) + H_{12}^*(\Delta + k\eta; p_k^0) \\
& + \frac{f(\Delta)}{f(-k\eta)} H_1^*(-k\eta)\Bigg] + o(n^{-1/2}),
\end{aligned} \tag{11'}
$$

where $p_k^0 = F(-k\eta)$.

Thus, $\hat{F}$ is $n^{-1/2}$ consistent on any given finite interval.

THEOREM 2. *Assume that* $\gamma \in (0, k_0\eta]$ *in* (9). *For any* $y' > \Lambda^{-1}(-\gamma)$ *and* $y'' < \Lambda^{-1}(1+\gamma)$, *we have*

$$\operatorname*{plim}_{n\to\infty} \sup_{y' \le y < y''} |\hat{\Lambda}(y) - \Lambda(y)| = 0.$$

Moreover,

$$n^{1/2}\big[\hat{\Lambda}(y) - \Lambda(y)\big] \Rightarrow_{\mathscr{L}} H_2(y)$$

over the interval (y', y''), *where* $H_2(y)$ *is a zero-mean tight Gaussian process. Here* $\hat{\Lambda}(y) - \Lambda(y)$ *has the representation*

$$\hat{\Lambda}(y) - \Lambda(y) = n^{-1/2}H_{21}^*(y) + n^{-1/2}H_{22}^*(y) + o(n^{-1/2}), \tag{12}$$

where

$$
\begin{aligned}
H_{12}^*(y) &= \frac{n^{-1/2}}{R(y)} \sum_{|z_i - \Lambda(y)| \le \gamma,\ z_i \in [0,1]} \frac{1(\varepsilon_i \le \Lambda(y) - z_i) - F(\Lambda(y) - z_i)}{f(\Lambda(y) - z_i)\, d(z_i)}, \\
H_{22}^*(y) &= -\int_{\max(0, \Lambda(y)-\gamma)}^{\min(1, \Lambda(y)+\gamma)} \frac{H_1(\Lambda(y) - z)}{f(\Lambda(y) - z)}\, dz \\
&= \int_{\min(\Lambda(y), \gamma)}^{\max(\Lambda(y)-1, -\gamma)} \frac{H_1(z)}{z}\, dz
\end{aligned}
$$

and $R(y) = \min(1, \Lambda(y) + \gamma) - \max(0, \Lambda(y) - \gamma)$ *satisfies the inequality* $R(y) \ge \min(\Lambda(y') + \gamma, 1 - (\Lambda(y'') - \gamma)) > 0$.

Therefore, if $[-c, c]$ is an interval such that $F(-c-\alpha) > 0$ and $F(c+\alpha) < 1$ for certain $\alpha > 0$, by choosing $\gamma = c$ in (9), the estimator $\hat{\Lambda}(y)$ is consistent on the interval $[\Lambda^{-1}(-c), \Lambda^{-1}(c)]$. Thus, if $f(\varepsilon) > 0$ for $\varepsilon \in (-\infty, +\infty)$, $\hat{\Lambda}$ is consistent on any finite interval in the support of Λ.

Again (12) consists of two additive parts. The first part is the error incurred by the estimation of $G(y \mid z)$, assuming that F is known. From the expression (12),

$$\operatorname{Var}(H_{21}^*(y)) \approx \frac{1}{R(y)} \int_{\max(0,\,\Lambda(y)-\gamma)}^{\max(1,\,\Lambda(y)+\gamma)} \frac{F(\Lambda(y)-z)(1-F(\Lambda(y)-z))}{f^2(\Lambda(y)-z)\,d(z)}\,dz$$
$$= \frac{1}{R(y)} \int_{\max(\Lambda(y)-1,\,-\gamma)}^{\min(\Lambda(y),\,\gamma)} \frac{F(z)(1-F(z))}{f^2(z)\,d(\Lambda(y)-z)}\,dz.$$

One can see that the variance of $\hat{\Lambda}$ does not depend on the derivative of Λ directly, in the sense that, if Z is uniformly distributed, $\operatorname{Var}(H_{21}^*(y))$ is constant over the range $\gamma \le \Lambda(y) \le 1-\gamma$. The second part, $H_{22}^*(y)$, which is the amount of accuracy we lose for not knowing F, is a Gaussian random process that is constant on $\gamma \le \Lambda(y) \le 1-\gamma$. This implies that not knowing F results in a random constant shift in the vertical direction in the estimate of $\Lambda(y)$ on this range. The magnitude of this shift depends on the weighted integral of $\hat{F} - F$. Although the covariance function of the overall error $H_2(y)$ is more complex, it is still true that it depends on Λ only through the density function of z. Therefore, if $d(z)$ is uniform, the variance of $\hat{\Lambda}$ is constant on the range $\Lambda^{-1}(\gamma) \le y \le \Lambda^{-1}(1-\gamma)$ and is independent of the value and derivative of Λ. This would imply that the variance of the prediction function $\hat{\Lambda}^{-1}$ is approximately proportional to the reciprocal of the derivative of Λ.

SKETCH OF PROOF OF THEOREM 1. The theorem is proved in several steps:

1. The discrepancy $\hat{F}(\Delta) - F(\Delta)$ can be decomposed into two major and some minor terms. The minor terms are smaller than $n^{-1/2}$ uniformly. The assumption that $d(z) > m_d > 0$ is necessary since the denominator $\#I(z)$ needs to go to infinity at a certain rate.
2. The first major part consists of the random variation around the true quantile curve, which is the shifted true median curve. The behavior of this component is the same as the ordinary empirical cumulative distribution and is completely unrelated to bandwidth (i.e., the number of intervals). Root-n consistency and asymptotic normality are straightforward to establish.
3. The second major part is related to the discrepancy between the true and estimated median curves. Since both curves run through the data set, one can infer that the total discrepancy can be expressed as the corresponding probability plus a random variation that is uniformly smaller than $n^{-1/2}$. More accurately, the random variation term is of order $n^{-3(1-r)/4}\ln n$. Thus, it is required that $r < 1/3$.
4. The total probability of the discrepancy in part 3 can be expressed as the area between the true and estimated median curves. A uniform Bahadur representation of the quantiles is obtained. The reminder term has order $n^{-3(1-r)/4}\ln n$, which is smaller than $n^{-1/2}$ for $r < 1/3$. The main component, which is $H_{12}^*(\Delta, p)$, is of order $n^{-1/2}$ and asymptotically normal. The

second-order differentiability is needed for the existence of the representation and to bound the minor terms, and the assumption that $r > 1/4$ makes sure that the interval approximation to the conditional distribution is accurate enough. If we use the shifted pth quantile estimator, the assumption that $f(F^{-1}(p)) > 0$ is necessary.

SKETCH OF PROOF OF THEOREM 2. Since the estimator involves $\hat{F}^{-1}(p)$, the condition that $f(\delta) > m_f > 0$ is a necessary one to ensure the uniqueness of $\hat{F}^{-1}(p)$. It is only necessary for δ such that $f(\delta) \in (-\gamma - c, \gamma + c)$ for a constant $c > 0$.

1. The assumption that $r > 1/4$ is necessary so that the approximation error of $G(y \mid I(z_0)) = \int_{z \in I(z_0)} G(y \mid z)\, dz$ to $G(y \mid z_0)$ is of order $o(n^{-1/2})$.
2. Another part of the error derives from the weight function, which uses $\hat{F}(\gamma)$ instead of $F(\gamma)$. This requires that F has no point mass at γ.
3. The main part of the discrepancy $\hat{\Lambda}(y) - \Lambda(y)$ is $\hat{F}^{-1}(\hat{p}(y)) - F^{-1}(p(y))$. This can be factored into two parts. The first part involves $\hat{p}(y) - p(y)$, and the second part involves $\hat{F}^{-1}(p) - F^{-1}(p)$. The second part is accounted for by Theorem 1. The first part, with the assumption that $r < 1/3$, is of order $O(n^{-1/2})$.

4. Implementations and simulations. In this section, we present simulation results and discuss some tricks in the practical use of the algorithm.

We assume that Z is a scalar with the standard normal $N(0,1)$ distribution. Three error distributions are considered:

1. $\varepsilon \sim N(0,1)$;
2. $\varepsilon \sim \text{Uniform}(-2,2)$;
3. $\varepsilon \sim$ Cauchy distribution with density $1/\pi(1+\varepsilon^2)$.

Strictly speaking, $\varepsilon \sim \text{Uniform}(-2,2)$ satisfies Assumption 4 only on the interval $(-2,2)$, but the procedure still works well for estimating F on a larger interval, too. The transformations we use are:

1. $\Lambda(y) = y$;
2. $\Lambda(y) = \log(y)$;
3. $\Lambda(y) = (1/13)\sinh(2y)$.

These transformation functions are also used in Horowitz (1996). The weighting function $w(p)$ for estimating Λ is defined to be

$$w(p) = \begin{cases} 1, & \text{if } p \in (0.1, 0.9); \\ 0, & \text{otherwise.} \end{cases}$$

For comparison purposes, we compute the estimates obtained by the median nonparametric regression (MNR) and the estimator by Horowitz (1996). Following Horowitz, let

$$K_2(x) = \tfrac{15}{16}(1 - x^2)^2 1(|x| \le 1),$$
$$K_6(x) = \tfrac{315}{2048}(15 - 140x^2 + 378x^4 - 396x^6 + 143x^8) 1(|x| \le 1),$$

where $1(x < y)$ is the indicator function for $x < y$. For MNR, $\lambda^{-1}(z)$ is estimated by the median of the following estimate of the conditional CDF of Y:

$$F(y \mid z) = [nhp(z)]^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_2\left(\frac{Z_i - z}{h}\right),$$

where

$$p(z) = (nh)^{-1} \sum_{i=1}^{n} K_2\left(\frac{Z_i - z}{h}\right),$$

and h is a bandwidth parameter. In the simulations, the bandwidth h minimizes the mean integrated squared error (MISE) of the median function Λ^{-1} defined later. Given two bandwidths h_z and h_y, let

$$\hat{G}(y \mid z) = [nh_z p_n(z)]^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_6\left(\frac{Z_i - z}{h_z}\right),$$
$$\hat{G}'_1(y \mid z) = [nh_z h_y p_n(z)]^{-1} \sum_{i=1}^{n} 1(Y_i \le y) K_2\left(\frac{Y_i - y}{h_y}\right) K_6\left(\frac{Z_i - z}{h_z}\right),$$
$$\hat{G}'_2(y \mid z) = \frac{\partial \hat{G}(y \mid z)}{\partial z}.$$

Horowitz's estimator for Λ is

$$\hat{\Lambda}(y) = -\int_{y_0}^{y} \int_{S_w} w(z) \frac{\hat{G}'_1(v \mid z)}{\hat{G}'_2(v \mid z)}\, dz\, dv,$$

where $y_0 = \Lambda^{-1}(-2)$ and S_w is the support for $w(z) = 0.5 \cdot 1(|z| < 1)$. The settings and the bandwidths used here are taken to be the same as in Horowitz (1996).

The unmodified estimator as described in the simulation section of Horowitz (1996) works very poorly. One reason is that $\hat{G}'_{nz}(v \mid z)$ is often close to 0, which blows up the estimator completely. The natural modification is to use a truncated version of it. Thus, in the following simulation, $\hat{G}'_{nz}(v \mid z)$ is replaced by $\hat{G}'^{*}_{nz}(v \mid z) = \min(\hat{G}'_{nz}(v \mid z), -d)$, where $d > 0$. Notice that, for Λ increasing, $G'_{nz}(v \mid z) < 0$. The value of d is taken to be 0.1 for $n = 100$ and 0.05 for $n = 1000$, since the minimal value of $\hat{G}'_{nz}(v \mid z)$ is about -0.3. Smaller values of d would yield worse results.

TABLE 1
Comparisons of MISE for $\hat{\Lambda}^{-1}$, $n = 100$ (*the SEs of the MISE based on* 1000 *simulations are given in parentheses*)

		Test function		
		Linear	**Log**	**Sinh**
Normal	Horowitz	0.56 (0.022)	3.89 (0.115)	1.67 (0.068)
	MNR	0.36 (0.007)	2.48 (0.063)	0.43 (0.013)
	SQE	0.22 (0.005)	1.59 (0.044)	0.35 (0.012)
	MSQE	0.20 (0.005)	1.48 (0.044)	0.29 (0.012)
	MLE	0.09 (0.003)	0.98 (0.052)	0.11 (0.004)
Uniform	Horowitz	0.83 (0.030)	5.03 (0.139)	2.68 (0.103)
	MNR	0.64 (0.012)	3.82 (0.101)	0.88 (0.028)
	SQE	0.33 (0.008)	2.25 (0.067)	0.60 (0.021)
	MSQE	0.34 (0.009)	2.21 (0.065)	0.59 (0.022)
Cauchy	Horowitz	0.71 (0.024)	5.33 (0.147)	3.81 (0.149)
	MNR	0.59 (0.004)	3.47 (0.131)	0.76 (0.026)
	SQE	0.51 (0.002)	2.76 (0.079)	0.74 (0.034)
	MSQE	0.48 (0.010)	2.48 (0.068)	0.57 (0.024)

Table 1 compares different methods for estimating Λ. The criterion for comparison is the mean integrated squared error

$$\mathrm{MISE}(\hat{\Lambda}) = \tfrac{1}{1000} \sum_{t=1}^{1000} \int_{-2}^{2} \left(\hat{\Lambda}_t^{-1}(z) - \Lambda^{-1}(z)\right)^2 dz,$$

where Λ_t^{-1} is the tth Monte Carlo replicate. This criterion puts equal weight for all $z \in (-2, 2)$. Note that the average is taken with respect to both the error ε and Z. The SQE (shifted quantile estimator) is the procedure proposed in Section 2. The MSQE is the modified version of SQE with median estimate $\hat{q}_{0.5}(z)$ in (6) obtained from MNR. It is an attempt to smooth the medians slightly before being used for estimating F. Maximum likelihood estimates are also obtained from the parametric model

$$\Lambda(y) = \frac{\Lambda_0(y) - \alpha}{\boldsymbol{\beta}} = z + \varepsilon$$

under the assumption of known Λ_0 and F and unknown α and $\boldsymbol{\beta}$. This is equivalent to knowing the transformation up to a location and a scale. Included in the parentheses are the standard error (SE) corresponding to each MISE. For SQE, the number of intervals is $\delta_n = 20$ for $n = 100$. We will discuss further the bandwidth selection later. In all cases, MISE is smaller for the SQE and MSQE than for MNR, which is, in turn, better than Horowitz's estimator. Although the new procedure with the medians from MNR works better than that with simple medians in most cases, the difference is usually not very large.

TABLE 2
Comparisons of MISE for $\hat{\Lambda}^{-1}$, $n = 1000$ (*the SEs of the MISE based on* 100 *simulations for Horowitz's estimator and* 1000 *simulations for all others are given in parentheses*)

		Test function		
		Linear	**Log**	**Sinh**
Normal	Horowitz	0.070 (0.0077)	1.480 (0.148)	0.166 (0.020)
	MNR	0.057 (0.0011)	0.581 (0.017)	0.064 (0.002)
	SQE	0.025 (0.0006)	0.233 (0.008)	0.039 (0.002)
	MSQE	0.023 (0.0005)	0.232 (0.007)	0.033 (0.001)
	MLE	0.009 (0.0003)	0.093 (0.004)	0.009 (0.0004)
Uniform	Horowitz	0.089 (0.010)	1.032 (0.116)	0.232 (0.030)
	MNR	0.112 (0.002)	1.009 (0.030)	0.153 (0.005)
	SQE	0.043 (0.001)	0.390 (0.014)	0.075 (0.003)
	MSQE	0.042 (0.001)	0.360 (0.012)	0.069 (0.003)
Cauchy	Horowitz	0.263 (0.033)	1.682 (0.193)	0.516 (0.067)
	MNR	0.088 (0.002)	1.093 (0.043)	0.106 (0.003)
	SQE	0.081 (0.002)	0.709 (0.022)	0.093 (0.003)
	MSQE	0.078 (0.002)	0.763 (0.024)	0.079 (0.003)

Table 2 gives the comparison for $n = 1000$. The number of replications is 100 for Horowitz's estimate, due to the long computation time it takes, and 1000 for the others. This factor partly explains the difference in the standard error in Table 2. For SQE, the number of intervals is $\delta_n = 50$. It is interesting to notice that the improvement of the new estimators over the MNR is more for $n = 1000$ than that for $n = 100$ in most cases. This verifies the theoretical result discussed in Theorem 2 on the rate of convergence of the estimators.

We now consider the bandwidth selection. Our theoretical results indicate that for δ_n in a reasonable range, the effect of bandwidth is only secondary. Table 3 gives the MISE($\hat{\Lambda}$) and MISE($\hat{F}$) for different values of δ_n for three representative cases: normal-linear, uniform-logarithmic and Cauchy-sinh. It can be seen that they are not sensitive to the choice of δ_n. An intuitive explanation for this is that the estimators proposed here are integral-type estimators. As indicated in (9), if δ_n is large, the quantile and percentiles

TABLE 3
MISE($\hat{\Lambda}$) *and MISE*($\hat{F}$) *for SQE for different values of* δ_n, $n = 100$

$\boldsymbol{\delta_n}$		**5**	**15**	**25**	**35**
Normal-linear	Λ	0.258	0.222	0.220	0.228
	F	0.0168	0.0138	0.0144	0.0160
Uniform-log	Λ	2.493	2.284	2.307	2.362
	F	0.0253	0.0208	0.0226	0.0249
Cauchy-sinh	Λ	0.819	0.732	0.745	0.819
	F	0.0319	0.0288	0.0317	0.0358

TABLE 4
Comparisons of MISE($\hat{F}$) for $n = 100$ based on 1000 simulations

Error	SQE	MSQE
Normal	0.0139 (0.0004)	0.0102 (0.00029)
Uniform	0.0197 (0.0005)	0.0163 (0.00048)
Cauchy	0.0297 (0.0010)	0.0189 (0.00056)

estimates are more unstable, but there are more intervals for summation; if δ_n is small, there are fewer intervals for summation but the quantile and percentiles estimates are more stable. The two factors balance out and thus the result is relatively robust against different choices of δ_n. One can see that the MISE($\hat{\Lambda}$) stays within 120% of the minimum for δ_n from below 10 to 30 in almost all cases. Therefore, we fix $\delta_n = 20$ for $n = 100$ and $\delta_n = 50$ for $n = 1000$ in Tables 1 and 2.

Tables 1 and 2 show that the MSQE may have better performance than the SQE in the estimation of Λ^{-1}. This improvement is more substantial for the estimation of F. The MSQE uses the median estimate $\hat{q}_{0.5}(z)$ obtained from median nonparametric regression with optimal bandwidth h. Table 4 compares the SQE and MSQE for $n = 100$, $\delta_n = 20$ and h taken to be optimal. The comparison is in terms of the MISE over the range $(-2.5, 2.5)$. Table 5 gives the MISE($\hat{F}$) for different values of h under the MSQE, again indicating that the choice of h is unimportant for $\hat{F}$ as it is in a reasonable range. Note that $\hat{F}$ is independent of Λ; thus, Λ does not appear in Tables 4 and 5.

We have not studied theoretically the case of multiple X-variables, that is, model (1) with unknown $\boldsymbol{\beta}$. One may infer that the root-n rate should still hold in this case since $\boldsymbol{\beta}$ can be estimated at a root-n rate, resulting in a discrepancy of order root n between $X\hat{\boldsymbol{\beta}}$ and $X\boldsymbol{\beta}$. Some simulations will be carried out here to study the effect of the estimation error in $\hat{\boldsymbol{\beta}}$ on the performance of our new procedures. We measure the performance using the empirical mean square error (EMSE) of the estimate median, that is,

$$\mathrm{EMSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\Lambda}^{-1}\left(x_i\hat{\boldsymbol{\beta}}\right) - \Lambda^{-1}(x_i\boldsymbol{\beta})\right)^2.$$

TABLE 5
Comparison of MISE for $\hat{F}$ using MSQE for different values of h, $n = 100$

h	0.5	0.75	1.0	1.25	1.5
Normal	0.0107	0.0101	0.0102	0.0108	0.0121
Uniform	0.0168	0.0162	0.0163	0.0170	0.0186
Cauchy	0.0204	0.0193	0.0189	0.0191	0.0197

Note that this criterion puts more weight in neighborhoods with more observations. The simulation model is

$$\frac{\Lambda(Y_i) - \alpha}{\boldsymbol{\beta}_1} = X_{i1} + \boldsymbol{\beta}_2 X_{i2} + \varepsilon_i.$$

We consider only the normal-logarithmic case. Four cases are simulated:

1. $\boldsymbol{\beta}_1 = 1$, $\boldsymbol{\beta}_1 = 0$, $X_{i1}, X_{i2} \sim N(0,1)$ are independent;
2. $\boldsymbol{\beta}_1 = 1$, $\boldsymbol{\beta}_2 = 0$, $X_{i1}, X_{i2} \sim N(0,1)$ with correlation $\rho = 0.5$;
3. $\boldsymbol{\beta}_1 = 1/\sqrt{2}$, $\boldsymbol{\beta}_2 = 1$, $X_{i1}, X_{i2} \sim N(0,1)$ are independent;
4. $\boldsymbol{\beta}_1 = 1/\sqrt{3}$, $\boldsymbol{\beta}_2 = 1$, $X_{i1}, X_{i2} \sim N(0,1)$ with correlation $\rho = 0.5$,

and $\alpha = 0$ for all cases. The coefficients are chosen so that $X\boldsymbol{\beta} \sim N(0,1)$. We estimate $\boldsymbol{\beta}$ by a simple average derivative estimator [ADE, Härdle and Stoker (1989)]:

$$\hat{\boldsymbol{\beta}} = \frac{1}{\Sigma w(x_i)} \sum_i w(x_i) \hat{\boldsymbol{\beta}}^i,$$

where $\hat{\boldsymbol{\beta}}^i$ is the regression coefficient using observations in the neighborhood of x_i with radius 1 and $w(x_i) = 1$ if the number of observations is more than 5 and 0 otherwise.

Table 6 compares the MSQE and MNR for $n = 100$, $\delta_n = 20$ and the bandwidth of MNR h is optimal. The MSQE still performs better than the MNR by a significant amount when $\boldsymbol{\beta}$ is unknown and has to be estimated, although the improvement is somewhat smaller than the improvement when $\boldsymbol{\beta}$ is known.

To sum up, the simulation shows that the new methods give better estimates of the unknown quantities and have a better rate of convergence than standard median nonparametric regression. The new method based on a stable median estimate gives the best result and is quite robust against choices of bandwidth parameters.

Figure 2 plots the true and estimated Λ for $n = 100$ for all nine combinations of Λ and F, with the 5–50–95% bands calculated from 1000 simulations. Results for Λ^{-1} and F are plotted in Figures 3 and 4. The estimates are quite close to the true underlying models.

TABLE 6
Comparisons of the EMSE for $n = 100$ based on 1000 simulations

Error	**MSQE**	**MNR**
Known $\boldsymbol{\beta}_2$	0.19 (0.006)	0.27 (0.007)
Case 1	0.27 (0.008)	0.33 (0.007)
Case 2	0.26 (0.007)	0.32 (0.007)
Case 3	0.29 (0.008)	0.35 (0.007)
Case 4	0.27 (0.007)	0.33 (0.008)

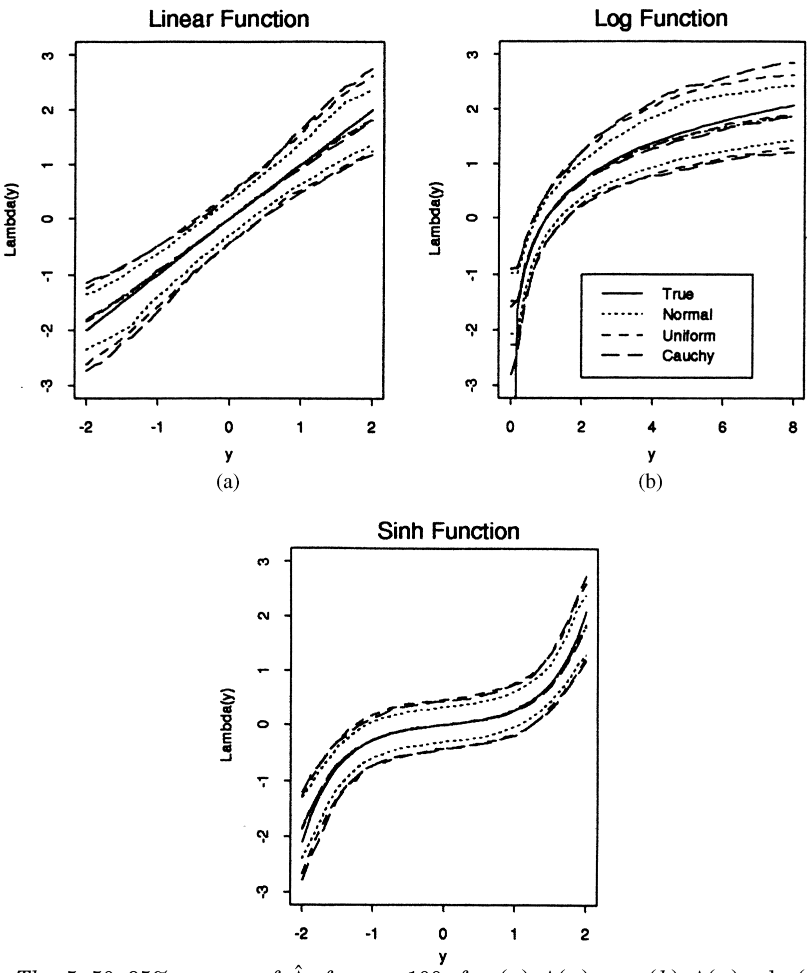

FIG. 2. *The* 5–50–95% *curves of* $\hat{\Lambda}$ *for* $n = 100$ *for* (a) $\Lambda(y) = y$, (b) $\Lambda(y) = \log(y)$ *and* (c) $\Lambda(y) = \sinh(2 * y)/13$, *each estimated under three types of error distributions. The line types are the same for all three plots.*

Another interesting result from Section 3 is that if the distribution of Z is uniform over its range, the variance of $\hat{\Lambda}$ is then constant over the range $\gamma \leq \Lambda(y) \leq 1 - \gamma$. To verify this, we take $\Lambda(y) = \log(y)$, $Z \sim U(-5, 5)$, $\varepsilon \sim N(0, 1)$ and $n = 1000$. The 5–50–95% bands of $\hat{\Lambda}(y) - \Lambda(y)$ are plotted in Figure 5 for $y \in [0.5, 8]$ and one can see that bands are quite stable, which confirms our theoretical result.

Horowitz (1996) uses only the normal case with $n = 100$ due to the heavy computation load of his estimators as he describes it. Our estimators require much less computation. Using FORTRAN 77, our estimators take about 0.1 second each simulation for $n = 100$ and 1–2 seconds for $n = 1000$, depending

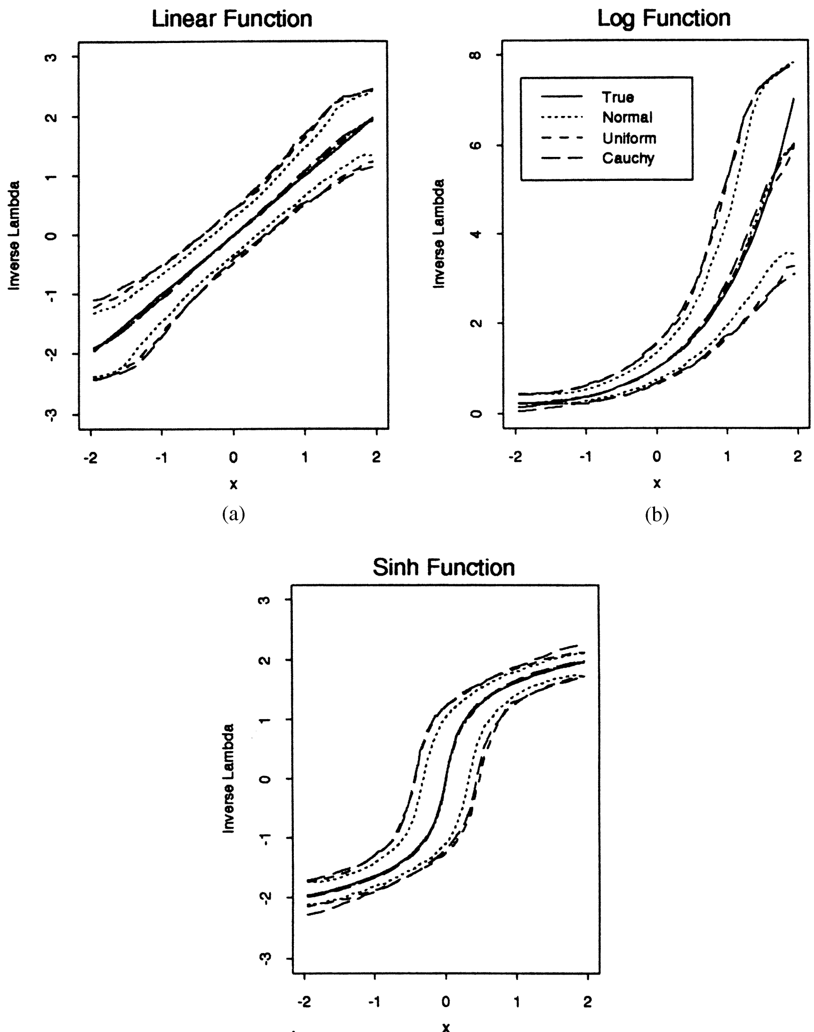

FIG. 3. *The 5–50–95% curves of* $\hat{\Lambda}^{-1}$ *for* $n = 100$ *for* (*a*) $\Lambda(y) = y$, (*b*) $\Lambda(y) = \log(y)$ *and* (*c*) $\Lambda(y) = \sinh(2 * y)/13$, *each estimated under three types of error distributions. The line types are the same for all three plots.*

on δ_n, on a SPARC 10/41. The relative computational savings could be even more significant for our estimators if one takes into account, the robustness of our estimators against different bandwidths.

5. Alternative algorithms. In this section, we propose several alternative methods. These alternative estimators also utilize the additivity property of the model (1), and thus resemble the one in Section 2 in spirit and may be even more efficient. Therefore, they should have similar behavior.

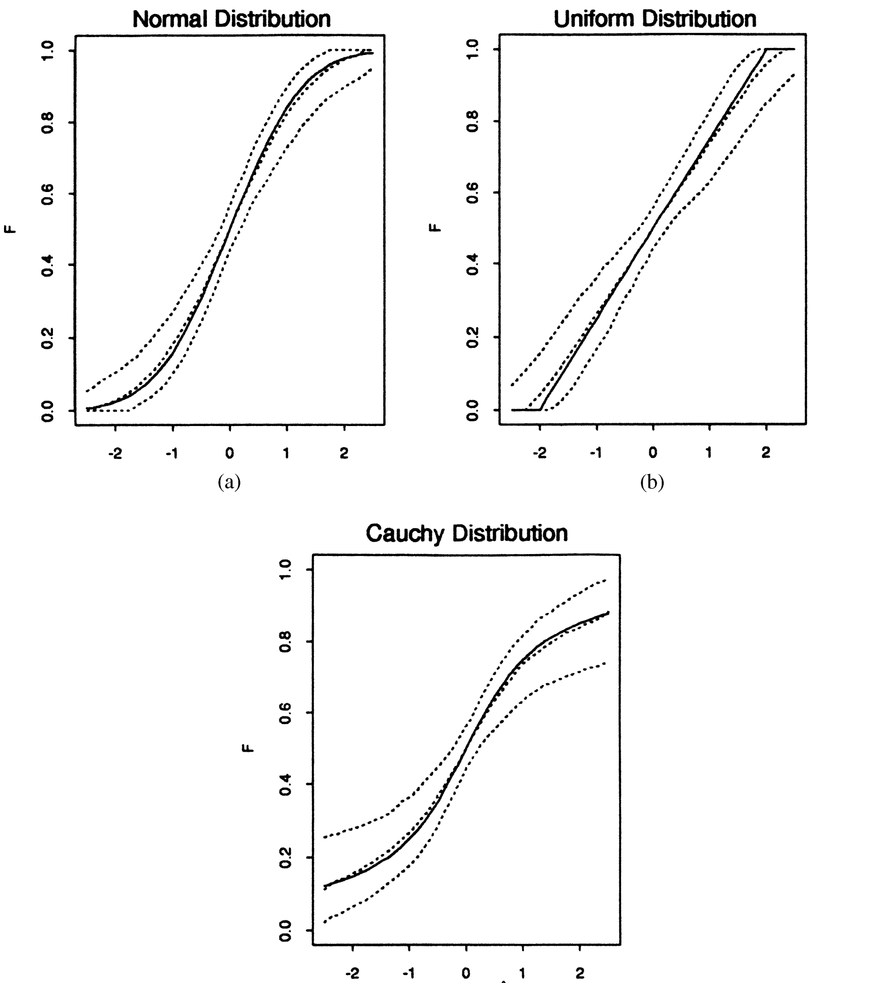

FIG. 4. *The true* F *and the* 5–50–95% *curves of* $\hat{F}$ *for* $n = 100$ *for* (*a*) $F \sim N(0,1)$, (*b*) $F \sim \text{Uniform}(-2,2)$ *and* (*c*) $F \sim \text{Cauchy}(0,1)$. $\hat{F}$ *is independent of* Λ.

Note that (1) implies

$$y = q_p(z) = \Lambda^{-1}(z + F^{-1}(p)). \tag{13}$$

Solving (13) for z, we have

$$q_p^{-1}(y) = \Lambda(y) - F^{-1}(p), \tag{14}$$

where $q_p^{-1}(y)$ is the value of z such that $G(y \mid z) = p$ for given y and p. Although theoretically $q_p^{-1}(y)$ is defined for all y and p, in the finite-sample case it is defined only on a limited range. Equation (14) is an additive function of $\Lambda(y)$ and $F^{-1}(p)$. Let $y_1^0, y_2^0, \dots, y_k^0$ be a grid of points in the

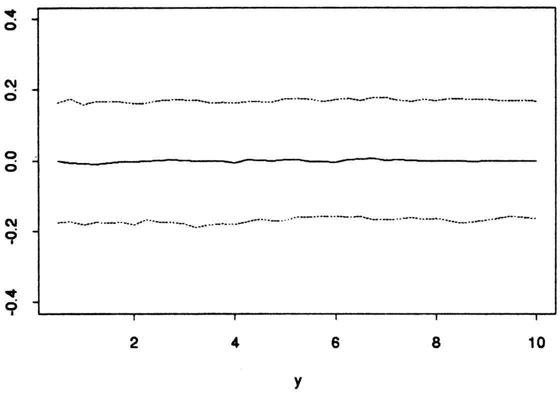

FIG. 5. *The* 5–50–95% *curves of* $\hat{\Lambda}(y) - \Lambda(y)$ *for* $\Lambda(y) = \log(y)$ *and* $Z \sim$ Uniform$(-5, 5)$ *and* $n = 1000$.

interval $(\min(y_j), \max(y_j))$ and let $p_1^0, p_2^0, \dots, p_m^0$ be a grid of points in the interval $[0, 1]$. Define

$$\hat{q}_p^{-1}(y_0) = \underset{z}{\operatorname{argmin}}\left(|\hat{p}(y \le y_o | I(z)) - p|\right),$$

where $\operatorname{argmin}_z(\cdot)$ is the value of z that minimizes the expression in the parentheses. Applying the standard (unbalanced) ANOVA, we can then obtain the estimates of $\Lambda(y_1^0), \Lambda(y_2^0), \dots, \Lambda(y_k^0)$ and $F^{-1}(p_1^0)$, $F^{-1}(p_2^0), \dots, F^{-1}(p_m^0)$ using the model

$$\hat{q}_{p_j^0}^{-1}\left(y_i^0\right) \cong \Lambda\left(y_i^0\right) - F^{-1}\left(p_j^0\right).$$

The difficulty of this algorithm lies in finding $\hat{q}_p^{-1}(y)$ for every y and p. To avoid this, we rearrange (3) and obtain

$$z = \Lambda\left(q_p(z)\right) - F^{-1}(p).$$

Substituting $q_p(z)$ by $\hat{q}(y \mid I(z))$ gives

$$z \approx \Lambda\left(\hat{q}_p(y \mid I(z))\right) - F^{-1}(p).$$

We can apply standard techniques for estimating additive models, possibly with an isotonization, to obtain $\hat{\Lambda}$ and $\hat{F}^{-1}$. To further reduce the computation load, we can restrict our attention to only a small set of p's in $(0, 1)$. A fitting procedure using a weight, such as $w(p) = p(1 - p)$, is often desirable to account for the heterogeneous variability in estimating $q_p(z)$.

Many possible extensions exist for the procedures proposed in this paper. The first and probably the easiest one is to censored data regression. Since the quantiles and percentiles can be estimated using Kaplan–Meier estimators if the data are censored, the extension is straightforward. Another possible extension is to the stationary semiparametric ARMA model. Although the estimators can be calculated exactly in the same way as in Section 2, the proof could be quite different.

Our remarks in Section 1 on the extension to the multiple regressor also apply to the model with the heteroscedastic error term $\varepsilon_i = (a + bx_i)\varepsilon_i'$, where $\varepsilon' \sim F$ and a and b are unknown parameters, since the parameters a and b can also be estimated at the root-n rate [see, e.g., Shen (1997)].

APPENDIX

Define

$$G(y \mid I(z)) = \frac{\int_{I(z)} G(y \mid z)\, d(z)\, dz}{\int_{I(z)} d(z)\, dz},$$

$$\hat{G}(y \mid I(z)) = \frac{\#\{(y_i < y,\, z_i \in I(z))\}}{\#\{I(z)\}},$$

where $I(z) = (z - 1/(2\delta_n), z + (1/2\delta_n))$ is the interval of width $1/\delta_n$ centered at z. Let $q_p(I(z))$ and $\hat{q}_p(I(z))$ be the pth quantile and sample quantile of y on the interval $I(z)$, that is,

$$q_p(I(z)) = \inf\{y; G(y \mid I(z)) \geq p\},$$

$$\hat{q}_p(I(z)) = \inf\{y; \hat{G}(y \mid I(z)) \geq p\}.$$

Let $1(\cdot)$ be an indicator function and, following Pollard (1984), we write $a_n \ll b_n$ for sequences $\{a_n\}$ and $\{b_n\}$ if $a_n/b_n \to 0$ as $n \to \infty$.

Proof of Theorem 1. The proof is for the shifted median estimator. The process is exactly the same for more general shifted pth quantile estimators. Reexpress $\hat{F}(-\Delta)$ into the following:

$$\begin{aligned}
&\hat{F}(-\Delta) - F(-\Delta) \\
&\quad = W_1 + W_2 + \frac{1}{(1-\Delta)\delta_n} \sum_{j=1}^{(1-\Delta)\delta_n} \frac{\sum_{z_i \in I(z_j^0 + \Delta)} \{1(\varepsilon_i \leq -\Delta) - P(\varepsilon \leq -\Delta)\}}{nd(z_i)/\delta_n} \\
&\quad = W_1 + W_2 + \frac{1}{(1-\Delta)n} \sum_{z_i \in (\Delta, 1)} \frac{1(\varepsilon_i \leq -\Delta) - P(\varepsilon \leq -\Delta)}{d(z_i)},
\end{aligned}$$

where

$$
\begin{aligned}
W_1 &= \frac{1}{(1-\Delta)\delta_n}\sum_{j=1}^{(1-\Delta)\delta_n}\frac{\sum_{z_i\in I(z_j^0+\Delta)}\left[1\left(y_i\le \hat{q}_{0.5}\left(I\left(z_j^0\right)\right)\right)-1\left(y_i\le q_{0.5}(z_i)\right)\right]}{\#\left\{I\left(z_j^0+\Delta\right)\right\}}\\
&= \frac{1}{(1-\Delta)\delta_n}\sum_{j=1}^{(1-\Delta)\delta_n}\frac{1}{\#\left\{I\left(z_j^0+\Delta\right)\right\}}\\
&\qquad\times\sum_{z_i\in I(z_j^0+\Delta)}\Big\{1\left(y_i\le \hat{q}_{0.5}\left(I\left(z_j^0\right)\right)\right)-1\left(y_i\le q_{0.5}(z_i)\right)\\
&\qquad\qquad-\Big[G\left(\hat{q}_{0.5}\left(I\left(z_j^0\right)\right)\,\middle|\, I\left(z_j^0+\Delta\right)\right)\\
&\qquad\qquad\quad-G\left(q_{0.5}\left(I\left(z_j^0\right)\right)\,\middle|\, I\left(z_j^0+\Delta\right)\right)\Big]\Big\}\\
&\quad+\frac{1}{(1-\Delta)\delta_n}\sum_{j=1}^{(1-\Delta)\delta_n}\Big\{G\left(\hat{q}_{0.5}\left(I\left(z_j^0\right)\right)\,\middle|\, I\left(z_j^0+\Delta\right)\right)\\
&\qquad\qquad-G\left(q_{0.5}\left(I\left(z_j^0\right)\right)\middle| I\left(z_j^0+\Delta\right)\right)\Big\}\\
&= W_{11}+W_{12}
\end{aligned}
$$

and

$$
\begin{aligned}
W_2 = \frac{1}{(1-\Delta)\delta_n}\sum_{j=1}^{(1-\Delta)\delta_n}\sum_{z_i\in I(z_j^0+\Delta)}&\left[\frac{1}{\#\left\{I\left(z_j^0+\Delta\right)\right\}}-\frac{1}{nd(z_i)/\delta_n}\right]\\
&\times\{1(\varepsilon_i\le -\Delta)-P(\varepsilon\le -\Delta)\}.
\end{aligned}
$$

Based on Lemmas 4–7, since $\Lambda(q_{0.5}^j(z_j^0)) = z_j^0$, we restate (11):

$$
\begin{aligned}
&\hat{F}(-\Delta)-F(-\Delta)\\
&\quad=\frac{1}{(1-\Delta)n}\sum_{z_i\in(\Delta,1)}\frac{1(\varepsilon_i\le -\Delta)-P(\varepsilon\le -\Delta)}{d(z_i)}\\
&\qquad-\frac{F(-\Delta)}{f(0)(1-\Delta)}\frac{1}{n}\sum_{z_i\in(0,1-\Delta)}\frac{1(\varepsilon_i\le 0)-0.5}{d(z_i)}+o(n^{-1/2}).
\end{aligned}\tag{11}
$$

It is worth noting here that $0\le\Delta\le\eta<1$ and that $d(z)\ge m_d>0$ from the assumptions. Notice also that if $\Delta=0$, we have $\hat{F}(0)-F(0)=o(n^{-1/2})$ from the preceding discussion, which is consistent with the initial condition that $\hat{F}(0)=F(0)=0.5$. To prove the uniform convergence of the first part of the right-hand side of (11), it is adequate to show the uniform convergence of

$$
n^{-1/2}\sum_{i=1}^{i=n}\frac{1(\varepsilon_i\le -\Delta)-P(\varepsilon\le -\Delta)}{d(z_i)}1(\Delta<z_i<1),
$$

which is true by Theorem 2.14 of Pollard (1984) if

$$h(z, \varepsilon; \Delta) = \frac{1(\varepsilon \le -\Delta) - P(\varepsilon \le -\Delta)}{d(z)} 1(\Delta < z_i < 1)$$

form an Euclidean class. Since the indicator function $1(\varepsilon \le -\Delta)$ and $I(\Delta < z_i < 1)$ both form Euclidean classes, by Lemma 2.14 of Pakes and Pollard (1989), $h(z, \varepsilon; \Delta)$ is Euclidean. The convergence thus follows. By Lemma 8, we then also have the convergence of the first part of the right-hand side of (11) to a Gaussian process. The convergence of the second part can be proved by the same argument.

Using the previous result, it is then straightforward to show that

$$n^{-1/2} \sum_{z_i \in (\Delta, 1)} \frac{1(\varepsilon_i \le -\Delta) - P(\varepsilon \le -\Delta)}{d(z_i)} \Rightarrow_{\mathscr{L}} H_{11}(-\Delta),$$

where $H_{11}(-\Delta)$ is a Gaussian process with zero mean and covariance function

$$R_{11}(-s, -t) = \operatorname{Cov}(H_{11}(-s), H_{11}(-t)) = F(-t)(1 - F(-s)) \int_t^1 \frac{1}{d(z)}\, dz,$$

where $0 \le s \le t \le \eta < 1$. Similarly,

$$n^{-1/2} \sum_{z_i \in (0, 1-\Delta)} \frac{1(\varepsilon_i \le 0) - 0.5}{d(z_i)} \Rightarrow_{\mathscr{L}} H_{12}(-\Delta),$$

where $H_{12}(-\Delta)$ is a Gaussian process with zero mean and covariance function

$$R_{12}(-s, -t) = \operatorname{Cov}(H_{12}(-s), H_{12}(-t)) = \frac{1}{4} \int_0^{1-t} \frac{1}{d(z)}\, dz,$$

where $0 \le s \le t \le \eta < 1$. The covariance between H_{11} and H_{12} is

$$\operatorname{Cov}(H_{11}(-s), H_{12}(-t)) = \begin{cases} 0.5\, F(-s) \int_t^{1-s} \frac{1}{d(z)}\, dz, & \text{for } t \le 1 - s, \\ 0, & \text{otherwise.} \end{cases}$$

Putting these together, we have

$$n^{1/2}(\hat{F}(-\Delta) - F(-\Delta)) \Rightarrow_{\mathscr{L}} H_1(-\Delta)$$

for $0 \le \Delta \le \eta$, where $H_1(-\Delta)$ is a zero-mean Gaussian process with covariance function (10). The proof is thus complete. □

PROOF OF COROLLARY 1. We will prove the result for $k = 1$. Recursive application of the same proof can be used to obtain the results for other k. Let $\hat{p}_1 = \hat{F}(\eta)$ and denote $\eta^* = F^{-1}(\hat{p}_1)$. By Theorem 1 with $p = \hat{p}_1$, we have

$$n^{1/2}\left[\hat{F}(\eta + \Delta) - F(\eta^* + \Delta)\right] = H_{11}^*(\Delta) + H_{12}^*(\Delta; \hat{p}_1) + o(n^{-1/2})$$

from (11), for $\Delta \in [-\eta, 0)$. Thus,

$$n^{1/2}\big[\hat{F}(\eta+\Delta) - F(\eta+\Delta)\big] = H_{11}^{*}(\Delta) + H_{12}^{*}(\Delta; \hat{p}_1) + \big[F(\eta^{*}+\Delta) - F(\eta+\Delta)\big] + o(n^{-1/2}).$$

By applying the Taylor expansion for $[F(\eta^{*}+\Delta) - F(\eta+\Delta)]$ and the uniform convergence of $\hat{F}(\Delta)$ over the range $[0, \eta + (1-\eta)/2]$, the desired result (11′) can thus be proved. □

PROOF OF THEOREM 2. Let $p_j(y) = G(y \mid I(z_j^0))$ and $\hat{p}_j(y) = \hat{G}(y|I(z_j^0))$.

$$\begin{aligned}\hat{\Lambda}(y) &= \sum_{j=1}^{\delta_n}\Big[z_j^0 + \hat{F}^{-1}\big(\hat{p}_j(y)\big)\Big]\frac{w\big(\hat{p}_j(y)\big)}{\Sigma w\big(\hat{p}_j(y)\big)}\\ &= \sum_{j=1}^{\delta_n}\Big[z_j^0 + F^{-1}\Big(G\big(y \mid z_j^0\big)\Big)\Big]\frac{w\big(\hat{p}_j(y)\big)}{\Sigma w\big(\hat{p}_j(y)\big)}\\ &\quad + \sum_{j=1}^{\delta_n}\Big[\hat{F}^{-1}\big(\hat{p}_j(y)\big) - F^{-1}\Big(G\big(y \mid z_j^0\big)\Big)\Big]\frac{w\big(\hat{p}_j(y)\big)}{\Sigma w\big(\hat{p}_j(y)\big)}.\end{aligned}$$

Since $\Lambda(y) = z + F^{-1}(G(y \mid z))$, we have

$$\hat{\Lambda}(y) = \Lambda(y) + U,$$

where

$$\begin{aligned}U &= \sum_{j=1}^{\delta_n}\Big[\hat{F}^{-1}\big(\hat{p}_j(y)\big) - F^{-1}\big(p_j(y)\big)\Big]\frac{w\big(\hat{p}_j(y)\big)}{\Sigma w\Big(\hat{p}\big(y, z_j^0\big)\Big)}\\ &\quad + \sum_{j=1}^{\delta_n}\Big[F^{-1}\big(p_j(y)\big) - F^{-1}\Big(G\big(y \mid z_j^0\big)\Big)\Big]\frac{w\big(\hat{p}_j(y)\big)}{\Sigma w\Big(\hat{p}\big(y, z_j^0\big)\Big)}\\ &= U_1 + U_2.\end{aligned}$$

Putting together Lemmas 9–11, after straightforward but tedious calculations similar to that in the proof of Theorem 1, and noting that $\Delta_j(y) = \Lambda(y) - z + o(n^{-r})$ for $z \in I(z_j^0)$, we restate (12):

$$\begin{aligned}&\hat{\Lambda}(y) - \Lambda(y)\\ &\quad = \sum_j \frac{w\big(p_j(y)\big)}{\Sigma_k w\big(p_k(y)\big)}\left\{\frac{\hat{p}_j(y) - p_j(y)}{f\big(\Delta_j(y)\big)} - n^{-1/2}\frac{H_1\big(\Delta_j(y)\big)}{f\big(\Delta_j(y)\big)}\right\}\\ &\qquad + o(n^{-1/2})\\ &\quad = \frac{1}{R(y)\,\delta_n}\sum_{|z_j^0 - \Lambda(y)| \le \gamma,\ j=1,\dots,\delta_n}\left\{\frac{\hat{p}_j(y) - p_j(y)}{f\big(\Delta_j(y)\big)}\right.\\ &\qquad\qquad \left. - n^{-1/2}\frac{H_1\big(\Delta_j(y)\big)}{f\big(\Delta_j(y)\big)}\right\} + o(n^{-1/2})\end{aligned} \tag{12}$$

$$= \frac{1}{nR(y)} \sum_{|z_i - \Lambda(y)| \le \gamma,\ z_i \in [0,1]} \frac{1(\varepsilon_i \le \Lambda(y) - z_i) - F(\Lambda(y) - z_i)}{f(\Lambda(y) - z_i)\, d(z_i)}$$

$$- n^{-1/2} \int_{\max(0,\Lambda(y)-\gamma)}^{\min(1,\Lambda(y)+\gamma)} \frac{H_1(\Lambda(y) - z)}{f(\Lambda(y) - z)}\, dz + o(n^{-1/2})$$

$$= V_1(y) + V_2(y) + o(n^{-1/2})$$

uniformly. The uniform convergence and the convergence of $n^{1/2}V_1(y)$ to a Gaussian process can be proved by showing that of the process

$$n^{-1/2} \sum_i \frac{1(\varepsilon_i \le \Lambda(y) - z_i) - F(\Lambda(y) - z_i)}{f(\Lambda(y) - z_i)\, d(z_i)} 1(|z_i - \Lambda(y)| \le \gamma),$$

which is immediate from Lemma 12, Theorem 2.14 of Pollard (1984) and Lemma 8. Letting $n^{1/2}V_1(y) \Rightarrow_{\mathscr{L}} H_{21}(y)$ and $0 < y_1 < y_2 < 1$, we have

$$\begin{aligned} R_{21}(y_1, y_2) &= \operatorname{Cov}(H_{21}(y_1), H_{21}(y_2)) \\ &= \frac{1}{R(y_1)R(y_2)} \\ &\quad \times \int_{\max(0,\Lambda(y_2)-\gamma)}^{\max(1,\Lambda(y_1)+\gamma)} \frac{F(\Lambda(y_1) - z)(1 - F(\Lambda(y_2) - z))}{f(\Lambda(y_1) - z) f(\Lambda(y_2) - z)\, d(z)}. \end{aligned}$$

On the other hand,

$$n^{1/2}V_2(y) = -\int_{\max(-\Lambda(y),-\gamma)}^{\min(1-\Lambda(y),\gamma)} \frac{H_1(-z)}{f(-z)}\, dz$$

is the integral in a Gaussian process and therefore is a Gaussian process itself, In fact, for y such that $\gamma < \Lambda(y) < 1 - \gamma$, we have

$$n^{1/2}V_2 = -\int_{-\gamma}^{\gamma} \frac{H_1(-z)}{f(-z)}\, dz,$$

which is a zero-mean Gaussian random variable independent of y with variance

$$R_{22} = \int_{-\gamma}^{\gamma} \int_{-\gamma}^{\gamma} \frac{R_1(-s, -t)}{f(-s) f(-t)}\, ds\, dt.$$

The proof is thus complete. □

LEMMA 1.

$$\left| \frac{\#\{I_n(z)\}}{n^{1-r}} - n^r \int_{I_n(z)} d(z)\, dz \right| \ll n^{-(1-r)/2} \ln n$$

uniformly for all $z \in [0, 1]$.

PROOF. Since $\#\{I_n(z)\} = \sum_i 1(z_i \in I_n(z))$, the function class $\mathscr{F}_n = 1(z_i \in I_n(z))$ is a Euclidean class [Pakes and Pollard (1989)], and

$$E1\big(z_i \in I_n(z)\big) = \int_{I_n(z)} d(z)\, dz \le M_d/\delta_n = Mn^{-r}$$

for some constant M. Thus, by Theorem 37 of Pollard (1984), we have

$$\frac{1}{n}\#\{I_n(z)\} - \int_{I_n(z)} d(z)\, dz \ll n^{-(1+r)/2}\ln n. \qquad \square$$

LEMMA 2.

$$\left|\hat{G}(y|I(z)) - G(y|I(z))\right| \ll n^{-(1-r)/2}\ln n$$

uniformly for all $z \in [0,1]$.

PROOF. Let $\mathscr{F}_n = \{1(y_i < y,\ z_i \in I(z));\ y \in (-\infty, +\infty),\ z \in [0,1]\}$. $\mathscr{F}_n$ is then a Euclidean class with

$$E1\big(y_i < y,\, z_i \in I(z)\big) = \int_{I(z)} G(y \mid z)\, d(z)\, dz \le M_d/\delta_n = Mn^{-r}$$

for some constant M. By Theorem 37 of Pollard (1984),

$$\frac{1}{n}\#\big\{\big(y_i < y,\, z_i \in I(z)\big)\big\} - \int_{I(z)} G(y \mid z)\, d(z)\, dz \ll n^{-(1+r)/2}\ln n$$

uniformly for all $z \in [0,1]$. From Lemma 1, we have

$$\left|\frac{c\#\{I(z)\}}{n^{1-r}\big(\delta_n \int_{I(z)} d(z)\, dz\big)} - 1\right| \to 0$$

uniformly, and $\delta_n \int_{I(z)} d(z)\, dz \ge m > 0$, the required result then follows. $\square$

LEMMA 3.

$\left|\hat{q}_p(I(z)) - q_p(I(z))\right| \ll n^{-(1-r)/2}\ln n$ *uniformly for* $z = z_j^0, j = 1, 2, \dots, \delta_n$.

PROOF. By definition, $G(y \mid I(z_j^0))$, $j = 1, \dots, n$, are continuous and have densities

$$g\big(y \mid I\big(z_j^0\big)\big) = \int_{I(z_j^0)} f(\Lambda(y) - z)\,\lambda(y)\, d(z)\, dz \Big/ \int_{I(z_j^0)} d(z)\, dz$$
$$= f(\Lambda(y) - z^*(y))\,\lambda(y)$$

by the mean value theorem, where $z^*(y) \in I(z_j^0)$. It is also easy to see that

$$q_p\left(1 - \frac{1}{2\delta_n}\right) < q_p(I(z)) < q_p\left(z + \frac{1}{2\delta_n}\right).$$

Thus,

$$m_f m_\lambda < f\Big(\Lambda\big(q_p(I(z)) - z^*\big(q_p(I(z))\big)\lambda\big(q_p(I(z)\big) < M_f M_\lambda$$

for n large enough. This shows that the conditional density $g(y \mid I(z_j^0))$ is bounded, which allows the application of Lemma 2.5.4B in Serfling (1980). Let $k_{n,j} = \#I(z_j^0)$ and

$$\varepsilon_{n,j} = \frac{4(\ln k_{n,j})^2}{m_f m_\lambda k_{n,j}}.$$

By an argument similar to that for Lemma 2.5.4B in Serfling (1980),

$$P\Big(|\hat{q}_p(I(z)) - q_p(I(z))| > \varepsilon_{n,j} \mid k_{n,j}\Big) \le \frac{2}{k_{n,j}^4}.$$

Let $k'_{n,j} = \frac{1}{2} n^{1-r} m_d$. We have

$$\begin{aligned}
&P\Big(\max_{j=1,\dots,\delta_n} \big|\hat{q}_p(I(z)) - q_p(I(z))\big| > \max_{j=1,\dots,\delta_n} (\varepsilon_{n,j})\Big) \\
&\quad \le \sum_{j=1,\dots,\delta_n} P\Big(\big|\hat{q}_p(I(z)) - q_p(I(z))\big| > \varepsilon_n\Big) \\
&\quad \le \sum_{j=1,\dots,\delta_n} \left(\frac{2}{k'^4_{n,j}} + P\Big(\#\big\{I\big(z_j^0\big)\big\} < k'_{n,j}\Big)\right) \\
&\quad \le \delta_n \frac{m_d^{-4(1-r)}}{n^{4(1-r)}} + \delta_n c_1 e^{-c_2 n}
\end{aligned}$$

for some constants c_1 and c_2. The last inequality follows from the Bernstein inequality [see Pollard (1984)]. The required result follows from the Borel–Cantelli lemma and the assumption that $r < 1/3$. □

LEMMA 4. $W_{11} = o(n^{-1/2})$ *uniformly.*

PROOF. For given positive constants c_1, c_2 and c_3, let

$$\begin{aligned}
&W_{11}\big(z_j^0, \Delta, m', m'', c'\big) \\
&\quad = \frac{1}{n^{1-r} c'} \sum_{z_i \in I(z_j^0 + \Delta)} \Big\{ I(y_i \le m') - I(y_i \le m'') \\
&\qquad\qquad - \Big[G\big(m' \mid I\big(z_j^0 + \Delta\big)\big) - G\big(m'' \mid I\big(z_j^0 + \Delta\big)\big)\Big]\Big\},
\end{aligned}$$

where m', m'' and c' are such that

$$|m' - m''| \le c_1 n^{-(1-r)/2} \ln n, \qquad \left|\frac{c'}{c_2} - 1\right| \le c_3 n^{-(1-r)/2} \ln n. \tag{15}$$

Let

$$d\big(z, y; z_j^0, \Delta, m', m'', c'\big) = I\big[z \in I\big(z_j^0 + \Delta\big)\big]\big[1(y \le m') - 1(y \le m'')\big]$$

and

$$\mathscr{F}_n = \Big\{ d\big(z, y; z_j^0, \Delta, m', m'', c'\big); j = 1, \dots, \delta_n; \\ \Delta \in \big(0, 1 - z_j^0\big); m', m'', c' \text{ satisfies } (15) \Big\}.$$

It is then straightforward to show that graphs of the function in $\mathscr{F}_n$ form a polynomial class of sets using Lemma 2.15 of Pollard (1984) and that

$$\begin{aligned} Ed\big(z, y; z_j^0, \Delta, m', m'', c'\big) &= \frac{1}{c'} \Big[G\big(m' \mid I\big(z_j^0 + \Delta\big)\big) - G\big(m'' \mid I\big(z_j^0 + \Delta\big)\big) \Big] \\ &\le \frac{2}{c_2} M |m' - m''| = O\big(n^{-(1-r)/2} \ln n\big) \end{aligned}$$

for n large enough and for some positive number M. Therefore, by Lemma 2.25 and Theorem 2.37 of Pollard (1984),

$$\sup_{\mathscr{F}_n} n^r \Big| W_{11}\big(z_j^0, \Delta, m', m'', c'\big) \Big| \ll n^{-3/4 - r/4} (\ln n)^2$$

almost surely. By Lemmas 1 and 3,

$$d\left(z, y; z_j^0, \Delta, \hat{q}_{0.5}\big(I\big(z_j^0\big)\big), q_{0.5}\big(I\big(z_j^0\big)\big), \frac{\#\big\{I\big(z_j^0 + \Delta\big)\big\}}{n^{1-r}} \right) \in \mathscr{F}_n$$

almost surely. Therefore,

$$\begin{aligned} &\sup_{j = 1, \dots, \delta_n; \Delta} \left| W_{11}\left(z_j^0, \Delta, \hat{q}_{0.5}\big(I\big(z_j^0\big)\big), q_{0.5}\big(I\big(z_j^0\big)\big), \frac{\#\big\{I\big(z_j^0 + \Delta\big)\big\}}{n^{1-r}} \right) \right| \\ &\qquad \ll n^{-1/2 - (1-3r)/4} (\ln n)^2 \ll n^{-1/2} \end{aligned}$$

uniformly over all $\Delta \in [0, 1]$ and $j = 1, 2, \delta_n$ with probability 1. The desired result then follows. □

LEMMA 5. $W_2 = o(n^{-1/2})$ *uniformly.*

PROOF. Reexpress W_2 to be

$$\begin{aligned} W_2 = {} & \frac{1}{(1-\Delta)\delta_n} \sum_{j=1}^{(1-\Delta)\delta_n} \left[\frac{1}{\#\big\{I\big(z_j^0 + \Delta\big)\big\}} - \frac{1}{n \int_{I(z_j^0)} d(z)\, dz} \right] \\ & \qquad \times \sum_{z_i \in I(z_j^0 + \Delta)} \{1(\varepsilon_i \le -\Delta) - P(\varepsilon \le -\Delta)\} \\ & + \frac{1}{(1-\Delta)\delta_n} \sum_{j=1}^{(1-\Delta)\delta_n} \sum_{z_i \in I(z_j^0 + \Delta)} \left[\frac{1}{n \int_{I(z_j^0)} d(z)\, dz} - \frac{1}{n d(z_j)/\delta_n} \right] \\ & \qquad \times \{1(\varepsilon_i \le -\Delta) - P(\varepsilon \le -\Delta)\} \\ = {} & W_{21} + W_{22}. \end{aligned}$$

Using Theorem 37 of Pollard (1984), it is straightforward to show that

$$\sup_{\Delta} \frac{1}{n}\left| \sum_{z_i \in I(z_j^0+\Delta)} \{1(\varepsilon_i \le -\Delta) - P(\varepsilon \le -\Delta)\}\right| \ll n^{-(1+r)/2}\ln n$$

almost surely. By Lemma 1, we also have

$$\sup_{j=1,\dots,\delta_n}\left| \frac{n}{\#\left\{I\left(z_j^0+\Delta\right)\right\}} - \frac{1}{\int_{I(z_j^0)} d(z)\,dz}\right| \ll n^{-(1+r)/2+2r}\ln n.$$

Therefore,

$$W_{21} \ll n^{-1+r}(\ln n)^2 \ll n^{-1/2}$$

uniformly. On the other hand,

$$\begin{aligned} W_{22} = \frac{1}{(1-\Delta)n}\sum_{i=1}^{n} I\left(z_i \in \left(z_1^0 + \Delta - \frac{1}{2\delta_n}, z^0_{(1-\Delta)\delta_n} + \Delta + \frac{1}{2\delta_n}\right)\right) \\ \times\left[\frac{1}{\delta_n \int_{I(z_i')} d(z)\,dz} - \frac{1}{d(z_j)}\right]\{1(\varepsilon_i \le -\Delta) - P(\varepsilon \le -\Delta)\}, \end{aligned}$$

where z_i' is the value of z_j^0 that minimizes $|z_j^0 - z_i|$, $j = 1,\dots,\delta_n$. Denote by $d(z,\varepsilon;\Delta)$ the function inside the summation sign. Since

$$\mathscr{F}_n' = \left\{\frac{1}{\delta_n \int_{I(z_i')} d(z)\,dz} - \frac{1}{d(z_j)}\right\}$$

contains only one function, it is obviously Euclidean class. Therefore, $\mathscr{F}_n = \{d(z,\varepsilon;\Delta)\}$ is an Euclidean class by Lemma 2.14 of Pakes and Pollard (1989), with

$$\begin{aligned} &Ed(z,\varepsilon;\Delta) \\ &\quad = F(-\Delta)(1-F(-\Delta)) \\ &\qquad \times \sum_{j=1}^{(1-\Delta)\delta_n} E\left\{\frac{1}{\delta_n \int_{I(z_j^0)} d(z)\,dz} - \frac{1}{d(z)}\right\}^2 I\left(z \in I\left(z_j^0\right)\right) \\ &\quad \le Mn^{-r} \end{aligned}$$

for some positive constant M. By Theorem 2.37 of Pollard (1984), we have

$$\sup_{\Delta\in(0,1)} |W_{22}| \ll n^{-(1+r)/2}\ln n \ll n^{-1/2}.$$

The desired result thus follows. □

LEMMA 6.

$$W_{12} = \frac{1}{(1-\Delta)\delta_n}\sum_j g\left(q_{0.5}^j \mid z_j^0 + \Delta\right)\left\{\hat{q}_{0.5}^j - q_{0.5}^j\right\} + \frac{1}{(1-\Delta)\delta_n}\sum_j D_n^j,$$

where $D_n^j \ll n^{-1/2}$ *uniformly with probability* 1.

PROOF. To avoid excessive notation, denote $\hat{q}^j_{0.5} = \hat{q}_{0.5}(I(z^0_j))$ and $q^j_{0.5} = q_{0.5}(I(z^0_j))$. Expanding $G(\hat{q}^j_{0.5} \mid z)$ about $q^j_{0.5}$ by second-order Taylor expansion, we have

$$\begin{aligned} G(\hat{q}^j_{0.5} \mid z) - G(q^j_{0.5} \mid z) &= \{\hat{q}^j_{0.5} - q^j_{0.5}\} \int_{I(z^0_j+\Delta)} g(q^j_{0.5} \mid z)\, d(z)\, dz \Big/ \int_{I(z^0_j+\Delta)} d(z)\, dz \\ &\quad + \{\hat{q}^j_{0.5} - q^j_{0.5}\}^2 \int_{I(z^0_j+\Delta)} g'_y(q^{j*}_{0.5} \mid z)\, d(z)\, dz \Big/ \int_{I(z^0_j+\Delta)} d(z)\, dz, \end{aligned}$$

where

$$g'_y(y_0 \mid z) = \frac{\partial g}{\partial y}\bigg|_{y=y_0} = f'(\Lambda(y_0) - z)\lambda^2(y_0) + f(\Lambda(y_0) - z)\lambda'(y_0)$$

and $q^{j*}_{0.5} \in (q^j_{0.5}, \hat{q}^j_{0.5})$. Substitute the first-order Taylor expansion of g about $z = z^0_j + \Delta$ into the first term on the right-hand side, and note that by Lemma 3 we have

$$G(\hat{q}^j_{0.5} \mid z) - G(q^j_{0.5} \mid z) = g(q^j_{0.5} \mid z^0_j + \Delta)\{\hat{q}^j_{0.5} + q^j_{0.5}\} + D^j_n,$$

where

$$\begin{aligned} D^j_n = \{\hat{q}^j_{0.5} - q^j_{0.5}\}^2 &\int_{I(z^0_j+\Delta)} g'_y(q^{j*}_{0.5} \mid z)\, d(z)\, dz \Big/ \int_{I(z^0_j+\Delta)} d(z)\, dz \\ &+ \{\hat{q}^j_{0.5} - q^j_{0.5}\} \int_{I(z^0_j+\Delta)} g'_z(q^{j*}_{0.5} \mid z)(z - z^0_j - \Delta)\, d(z)\, dz \Big/ \int_{I(z^0_j+\Delta)} d(z)\, dz. \end{aligned}$$

By Lemma 3 and Assumptions 3 and 4, we have

$$\sup_{\Delta, j=1,\dots,\delta_n} |D^j_n| = O(n^{\max(-(1+r)/2,\, r-1)} \ln n) \ll n^{-1/2}$$

almost surely for n large enough. □

LEMMA 7.

$$\hat{q}^j_{0.5} - q^j_{0.5} = -\frac{\hat{p}(q^j_{0.5}) - 0.5}{g(q^j_{0.5} \mid I(z^0_j))} + R_{nj},$$

where $R_{nj} \ll O(n^{-3(1-r)/4} \ln n)$ *uniformly. Furthermore,*

$$\hat{q}^j_{0.5} - q^j_{0.5} = -\sum_{z_i \in I(z^0_j)} \frac{1(\varepsilon_i \le 0) - 0.5}{n^{1-r} d(z_i) g(q^j_{0.5} \mid z_i)} + R'_{nj},$$

where again $R'_{nj} \ll O(n^{-3(1-r)/4} \ln n) \ll n^{-1/2}$.

PROOF. The proof of the first part can be carried out in the same way as the original Bahadur representation [see Bahadur (1966) or Serfling (1980) for references], with slight modifications to account for the uniformity across $j = 1, 2, \dots, \delta_n$ in the same way as in Lemma 3.

The second equality can be obtained from the first equality by bounding the reminder terms, using an argument similar to that for Lemma 5. □

LEMMA 8. *If the set of functions* $\mathscr{F}_n = h(z, y; \Delta)$, $\Delta \in (0,1)$, *forms a Euclidean class and* $\{(z_i, y_i), i = 1, 2, \dots\}$, *are independent realizations from the independent random vector* (Z, Y), *then*

$$H_\Delta = \frac{1}{n^{1/2}} \sum_{i=1}^{i=n} h(z_i, y_i; \Delta)$$

converges to a Gaussian process indexed by Δ.

PROOF. See part b of the proof of Theorem 1 in Horowitz (1996), utilizing Lemma 2.16 of Pakes and Pollard (1989). □

LEMMA 9. $U_2 \ll n^{-1/2}$ *uniformly with probability* 1.

PROOF. After straightforward calculation using second-order Taylor expansion, we have

$$\hat{p}_j(y) - G\big(y \mid z_j^0\big) = O(n^{-2r})$$

uniformly for all y and j. In view of Lemma 2, there exists a positive constant η with probability 1 so that, if $w(\hat{p}_j(y)) = 1$,

$$p_j(y), G\big(y \mid z_j^0\big) \in \big(\hat{F}(\gamma) - \eta, \hat{F}(\gamma) + \eta\big).$$

Therefore, we have

$$\Big|F^{-1}\big(p_j(y)\big) - F^{-1}\Big(G\big(y \mid z_j^0\big)\Big)\Big| w\big(\hat{p}_j(y)\big) = (n^{-2r}) \ll n^{-1/2}$$

uniformly for all y and j. The desired result thus follows. □

LEMMA 10. *With probability* 1, *for all j such that* $w(\hat{p}_j(y)) = 1$, *we have*

$$\begin{aligned}
&\hat{F}^{-1}\big(\hat{p}_j(y)\big) - F^{-1}\big(p_j(y)\big)\\
&\qquad = \frac{1}{f\big(\Delta_j(y)\big)}\Big\{\hat{p}_j(y) - p_j(y) - n^{-1/2}H_1\big(\Delta_j(y)\big)\Big\} + o(n^{-1/2})
\end{aligned}$$

uniformly, where $\Delta'_j(y) = \hat{F}^{-1}(\hat{p}_j(y))$, $\Delta_j(y) = F^{-1}(p_j(y))$. *The same results hold for* $w(p_j(y)) = 1$.

PROOF. If $w(\hat{p}_j(y)) = 1$, $\Delta'_j(y) \in (-\gamma, \gamma)$. From the definition of $\Delta'_j(y)$ and $\Delta_j(y)$, we have

$$F(\Delta'_j(y)) - F(\Delta_j(y)) = \hat{p}_j(y) - p_j(y) - \big(\hat{F}(\Delta'_j(y)) - F(\Delta'_j(y))\big).$$

In view of Lemma 3 and Theorem 1 and the preceding equation,

$$F(\Delta'_j(y)) - F(\Delta_j(y)) = O(n^{-(1-r)/2}\ln n)$$

uniformly almost surely. Furthermore, using Assumption 2(a), with probability 1, if $w(\hat{p}_j(y)) = 1$, there exists a positive constant c_1, $\Delta_j(y) \in (-\gamma - c_1, \gamma + c_1)$ uniformly over all j and y for n large enough, and

$$\Delta'_j(y) - \Delta_j(y) = O(n^{-(1-r)/2}\ln n)$$

uniformly. By Theorem 1 and the equicontinuity property of the Gaussian processes, we have

$$F(\Delta'_j(y)) - F(\Delta_j(y)) = \hat{p}_j(y) - p_j(y) - n^{-1/2}H_1(\Delta_j(y)) + o(n^{-1/2})$$

uniformly for n large enough. It is thus straightforward that, with probability 1,

$$\Delta'_j(y) - \Delta_j(y) = \frac{1}{f(\Delta_j(y))}\big\{\hat{p}_j(y) - p_j(y) - n^{-1/2}H_1(\Delta_j(y))\big\} + o(n^{-1/2})$$

holds uniformly for all y and j such that $w(\hat{p}_j(y)) = 1$. □

LEMMA 11. *Let*

$$w(p(y)) = \begin{cases} 1, & \textit{if } p \in (F(-\gamma), F(\gamma)), \\ 0, & \textit{otherwise}. \end{cases}$$

Then $\sum_j |w(\hat{p}_j(y)) - w(p_j(y))| = O(\ln n)$. *Furthermore, with probability* 1, *there exists a constant* $0 < R(y) < +\infty$ *such that*

$$\sum_j w(p_j(y)) = R(y)\delta_n + O(1)$$

uniformly for all $y \in [\Lambda^{-1}(a), \Lambda^{-1}(b)]$.

PROOF. Assume that $\hat{F}(-\gamma) < \hat{F}(\gamma)$, since it holds with probability 1. For $w(\hat{p}_j(y)) - w(p_j(y)) \neq 0$ to be true, either $\hat{p}_j(y)$ or $p_j(y)$ must be in one of the following sets:

$$\big(\hat{F}(-\gamma), F(-\gamma)\big), \quad \big(F(-\gamma), \hat{F}(-\gamma)\big), \quad \big(\hat{F}(\gamma), F(\gamma)\big), \quad \big(F(\gamma), \hat{F}(\gamma)\big),$$

while the other must not be.

From Theorem 1 and Lemma 2, there exist positive constants c_1 and c_2, so that, with probability 1,

$$|\hat{F}(\Delta) - F(\Delta)| < c_1 n^{-1/2} \ln n$$

for $\Delta = \pm\gamma$ and

$$|\hat{p}_j(y) - p_j(y)| < c_2 n^{-(1-r)/2} \ln n.$$

By the mean value theorem, we have

$$p_j(y) = G\big(y \mid I(z_j^0)\big) \in \left(G\left(y \,\middle|\, z_j^0 - \frac{1}{2\delta_n}\right), G\left(y \,\middle|\, z_j^0 + \frac{1}{2\delta_n}\right)\right),$$

which implies that there exists a constant c_3, so that

$$|G(y \mid z) - p_j(y)| < c_3 n^{-r}.$$

Note that $r < (1-r)/2$ for $r < 1/3$. Putting together the previous inequalities, there exists a constant c_4, so that, for $w(\hat{p}_j(y)) - w(p_j(y)) \neq 0$ to be true, with probability 1,

$$F(\Delta) - cn^{-r} \ln n < G(y \mid z) < F(\Delta) + c_4 n^{-r} \ln n$$

for either $\Delta = -\gamma$ or $\Delta = \gamma$ uniformly for y. By the assumptions, there exists a c_5, so that

$$z \in \big(\Lambda(y) + \Delta - c_5 n^{-r} \ln n, \Lambda(y) + \Delta + c_5 n^{-r} \ln n\big)$$

for $\Delta = \pm\gamma$, which contains at most $2(c_5 n^{-r} \ln n)\delta_n = O(\ln n)$ of intervals of length $1/\delta_n$.

Following the same line of arguments, there exists a constant c_6, such that for every z_j^0 such that

$$\max\big(0, \Lambda(y) - \gamma + c_6 n^{-r}\big) < z_j^0 < \min\big(1, \Lambda(y) + \gamma - c_6 n^{-r}\big),$$

we have $w(p_j(y)) = 1$. Thus, there exists a constant $0 < R(y) < +\infty$, such that

$$\sum_j w\big(p_j(y)\big) = R(y)\delta_n + O(1)$$

uniformly for all y, where

$$\begin{aligned} R(y) &= \min(1, \Lambda(y) + \gamma) - \max(0, \Lambda(y) - \gamma) \\ &= \begin{cases} \min(1, \Lambda(y) + \gamma), & \text{if } \Lambda(y) - \gamma < 0, \\ 2\gamma, & \text{if } -\gamma < \Lambda(y) \leq \gamma, \\ 1 - \max(0, \Lambda(y) - \gamma), & \text{if } \Lambda(y) + \gamma > 1, \end{cases} \end{aligned}$$

using the fact that $\delta_n = [cn^r]$. It is obvious that $R(y)$ is an equicontinuous function of y. $\square$

LEMMA 12. *The class of functions*

$$h(z, \varepsilon; T) = \frac{1(\varepsilon \le T - z) - F(T - z)}{f(T - z)\, d(z)} 1(|z - T| \le \gamma)$$

for $0 \le T \le 1$ *is Euclidean.*

PROOF. The result follows immediately from Lemmas 2.13 and 2.14 of Pakes and Pollard (1989) using the facts that both $1(\varepsilon \le T - z)$ and $1(|z - T| \le \gamma)$ are Euclidean and T has bounded range. □

Acknowledgments. The authors thank Joel L. Horowitz for motivating our interests in this topic and for making his stimulating preprint and simulation program available to us, and Peter Bickel for encouraging us to study Horowitz's nonparametric transformation estimate. We would also like to thank Ruey S. Tsay, the editors and a referee for giving us many helpful comments.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BICKEL, P. J. and DOKSUM, K. A. (1981). An analysis of transformations revisited. *J. Amer. Statist. Assoc.* **76** 296–311.

BOX, G. E. P. and COX, D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser B* **26** 211–252.

BOX, G. E. P. and COX, D. R. (1982). Comment on "An analysis of transformations revisited," by P. J. Bickel and K. A. Doksum. *J. Amer. Statist. Assoc.* **77** 209–210.

BREIMAN, L. and FRIEDMAN, J. H. (1985). Estimating optimal transformations for multiple regression and correlation, *J. Amer. Statist. Assoc.* **80** 580–598.

CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, London.

CHAUDHURI, P., DOKSUM, K. and SAMAROV, A. (1994). Nonparametric estimation of global functionals based on quantile regression. Unpublished manuscript.

DOKSUM, K. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15** 325–345.

DUAN, N. (1990). The adjoint projection pursuit regression. *J. Amer. Statist. Assoc.* **85** 1029–1038.

EFRON, B. (1982). Transformation theory: how normal is a family of distributions? *Ann. Statist.* **10** 323–339.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

HÄRDLE, W. and STOKER, T. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995.

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

HINKLEY, D. V. and RUNGER, G. (1984). The analysis of transformed data (with discussion). *J. Amer. Statist. Assoc.* **79** 302–320.

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

KRUSKAL, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *J. Roy. Statist. Soc. Ser. B* **27** 251–263.

MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22** 712–731.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23** 182–198.

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

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

SHEN, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591.

WANG, N. and RUPPERT, D. (1995). Nonparametric estimation of the transformation in the transformation-both-sides regression model. *J. Amer. Statist. Assoc.* **90** 522–534.

WANG, N. and RUPPERT, D. (1996). Estimation of regression parameters in a semiparametric transformation model. *J. Statist. Plann. Inference* **52** 331–351.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637
E-MAIL: jmy@gsbjmy.uchicago.edu

STATISTICS GROUP
RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CALIFORNIA 90401
E-MAIL: naihua_duan@rand.org