The Annals of Statistics
1993, Vol. 21, No. 4, 1900–1925

NONPARAMETRIC REGRESSION WITH ERRORS IN VARIABLES

BY JIANQING FAN AND YOUNG K. TRUONG

University of North Carolina

The effect of errors in variables in nonparametric regression estimation is examined. To account for errors in covariates, deconvolution is involved in the construction of a new class of kernel estimators. It is shown that *optimal* local and global rates of convergence of these kernel estimators can be characterized by the tail behavior of the characteristic function of the error distribution. In fact, there are two types of rates of convergence according to whether the error is ordinary smooth or super smooth. It is also shown that these results hold uniformly over a class of joint distributions of the response and the covariate, which is rich enough for many practical applications. Furthermore, to achieve optimality, we show that the convergence rates of all possible estimators have a lower bound possessed by the kernel estimators.

1. Introduction. A tremendous amount of attention has been focused on the problem of nonparametric regression estimation. Most of this attention has been directed to data with standard structure. On the other hand, regression analysis with errors in variables is evolving rapidly. See, for example, Anderson (1984), Carroll, Spiegelman, Lan, Bailey and Abbott (1984), Stefanski (1985), Stefanski and Carroll (1985, 1987), Fuller (1987), Prentice (1986), Bickel and Ritov (1987), Schafer (1987), Whittemore and Keller (1988) and Whittemore (1989). However, the latter has centered around the parametric approach in which the regression function is assumed to take on a particular functional form. The desire to examine the effect of errors in variables in nonparametric regression leads to the subject of this paper. Recently, Stefanski and Carroll (1991) give an interesting account on nonparametric calibration, which is also useful for errors-in-variables modeling.

Let (X, Z) denote a pair of random variables and consider the problem of estimating the regression function $m(x) = E(Z|X = x)$. Due to the measuring mechanism or the nature of the environment, the variable X is measured with error and is not directly observable [Fuller (1987), page 2]. Instead, X is observed through $Y = X + \varepsilon$, where ε is a random disturbance. To make this nonparametric problem identifiable, it is assumed that ε has a known distribution, and is independent of (X, Z). Given a random sample $(Y_1, Z_1), \dots, (Y_n, Z_n)$ from the distribution of (Y, Z), three interesting issues arise naturally: (a) How can a *nonparametric* regression function estimator be constructed to reflect the fact that there are errors in variables? (b) What can be said about its sampling behaviors? (c) Does it possess any optimal properties? The discussions of these issues form the core of the paper.

Received April 1990; revised November 1992.

AMS 1991 *subject classifications*. Primary 62G20; secondary 62G05, 62J99.

Key words and phrases. Errors in variables, nonparametric regression, deconvolution, kernel estimator, optimal rates of convergence.

The first issue is addressed by considering the following estimator:

$$\hat{m}(x) = \sum_j W_{n,j}(Y_1, \dots, Y_n) Z_j,$$

with weights $W_{n,j}(Y_1, \dots, Y_n)$ constructed to account for measurement errors. The basic idea is closely related to the deconvolution techniques, and details about this construction are given in Section 2.

The second issue is addressed by examining to what extent the distribution of ε affects the rates of convergence of the preceding nonparametric estimator, both locally and globally. These results depend on the smoothness of the error distribution. A distribution is called *ordinary smooth* if the tails of its characteristic function decay to 0 at an algebraic rate. It is called *super smooth* if its characteristic function has tails approaching 0 exponentially fast. For example, distributions such as double exponential and gamma are ordinary smooth, whereas normal and Cauchy are super smooth. Section 3 contains a formal definition of the smoothness of distributions.

Depending on the type of error distribution, the rates of convergence of the kernel estimators are quite different—the local and global rates are slower in the super smooth model, whereas they are faster in the ordinary smooth model. These results also hold uniformly over a class of joint distributions of (X, Z) which is rich enough for theoretical and practical applications. A detailed discussion is given in Section 3.

The third issue is rate optimality. This concerns the construction of minimax lower bounds for the rates of convergence. Indeed, Section 4 shows that these provide lower bounds for all possible estimators when the covariates are measured with errors. These results hold locally and globally, as well as uniformly over the aforementioned class of joint distributions of (X, Z). Moreover, the effects of the smoothness of the error distribution on rates of convergence are clearly demonstrated.

The effect of error distribution on the sampling properties of the proposed estimators is further examined by examples based on simulation. This is discussed in detail in Section 5. Concluding remarks are given in Section 6, and proofs are presented in Section 7.

2. Kernel estimators. Let $(X_1, Z_1), \dots, (X_n, Z_n)$ denote a random sample from the distribution of (X, Z) and let $K(\cdot)$ denote a kernel function. Recall that in the case that X is *observable*, the kernel estimator of the regression function $E(Z|X = x)$ is obtained by averaging the Z's with weights proportional to $K((x - X_j)/h_n)$:

$$\begin{aligned} \hat{m}_n(x) &= \sum_j K\left(\frac{x - X_j}{h_n}\right) Z_j \Big/ \sum_j K\left(\frac{x - X_i}{h_n}\right) \\ &= \frac{1}{nh_n} \sum_j K\left(\frac{x - X_j}{h_n}\right) Z_j / \hat{f}_n(x), \end{aligned} \tag{2.1}$$

where h_n is a smoothing parameter and $\hat{f}_n(x) = (nh_n)^{-1}\Sigma_i K((x - X_i)/h_n)$ is a kernel estimator of the density of covariate X.

Since the variables $X_1, \dots, X_n$ are *not observable*, the kernel estimator $\hat{f}_n(x)$ will be constructed from $Y_j = X_j + \varepsilon_j$, $j = 1, \dots, n$. Denote the densities of Y and X by $f_Y(\cdot)$ and $f_X(\cdot)$, respectively. Let $F_\varepsilon(\cdot)$ denote the distribution function of ε. Then

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(y - x)\, dF_\varepsilon(x).$$

This suggests that the marginal density function $f_X(\cdot)$ can be estimated by the deconvolution method. Using a kernel function $K(\cdot)$ with a bandwidth h_n, Stefanski and Carroll (1990) consider the following estimator:

$$\hat{f}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)}\,dt, \tag{2.2}$$

where $\phi_K(\cdot)$ is the Fourier transform of the kernel function $K(\cdot)$, $\phi_\varepsilon(\cdot)$ is the characteristic function of the error variable ε and $\hat{\phi}_n(\cdot)$ is the empirical characteristic function:

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_1^n \exp(itY_j).$$

See also Carroll and Hall (1988), Fan (1991a, 1991b), Liu and Taylor (1989) and Zhang (1990) for recent contributions in this area.

Note that (2.2) can be rewritten in the kernel form

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_1^n K_n\left(\frac{x - Y_j}{h_n}\right), \tag{2.3}$$

with

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\,dt. \tag{2.4}$$

Appealing to (2.1), (2.3) and (2.4), we propose the following kernel regression function estimator involving errors in variables:

$$\begin{aligned} \hat{m}_n(x) &= \sum_j K_n\left(\frac{x - Y_j}{h_n}\right) Z_j \Big/ \sum_i K_n\left(\frac{x - Y_i}{h_n}\right) \\ &= \frac{1}{nh_n}\sum_j K_n\left(\frac{x - Y_j}{h_n}\right) Z_j / \hat{f}_n(x), \end{aligned} \tag{2.5}$$

where $\hat{f}_n(x)$ and $K_n(x)$ are defined by (2.3) and (2.4). This estimator will be shown to possess many interesting asymptotic properties, which are the topic of the next two sections.

3. Performance of kernel estimators. The sampling behaviors of the kernel estimators (2.5) considered in the previous section will be treated here. The rates of convergence of these estimators depend on the smoothness of error distributions, which can be classified into the following:

1. Super smooth of order β: If the characteristic function of the error distribution $\phi_\varepsilon(\cdot)$ satisfies

$$d_0|t|^{\beta_0}\exp\left(-|t|^\beta/\gamma\right) \le |\phi_\varepsilon(t)| \le d_1|t|^{\beta_1}\exp\left(-|t|^\beta/\gamma\right) \quad \text{as } t \to \infty, \tag{3.1}$$

   where d_0, d_1, β and γ are positive constants and β_0 and β_1 are constants.
2. Ordinary smooth of order β: If the characteristic function of the error distribution $\phi_\varepsilon(\cdot)$ satisfies

$$d_0|t|^{-\beta} \le |\phi_\varepsilon(t)| \le d_1|t|^{-\beta} \quad \text{as } t \to \infty, \tag{3.2}$$

   for positive constants d_0, d_1 and β.

For example,

$$\text{Super smooth distributions:} \quad \begin{cases} N(0,1), & \text{with } \beta = 2, \\ \dfrac{1}{\pi}\dfrac{1}{1+x^2} \text{ Cauchy}(0,1), & \text{with } \beta = 1. \end{cases}$$

$$\text{Ordinary smooth distributions:} \quad \begin{cases} \dfrac{\alpha^p}{\Gamma(p)}x^{p-1}e^{-\alpha x} \text{ (gamma)}, & \text{with } \beta = p, \\ \frac{1}{2}e^{-|x|} \text{ (double exponential)}, & \text{with } \beta = 2. \end{cases}$$

The rates of convergence depend on β, the order of smoothness of the error distribution. They also depend on the smoothness of the regression function $m(x)$ and regularity conditions on the marginal distribution. Specifically, these conditions are as follows.

CONDITION 1.

(i) The characteristic function of the error distribution $\phi_\varepsilon(\cdot)$ does not vanish.
(ii) Let $a < b$. The marginal density $f_X(\cdot)$ of the unobserved X is bounded away from 0 on the interval $[a, b]$, and has a bounded kth derivative.
(iii) The regression function $m(\cdot)$ has a continuous kth derivative on $[a, b]$.
(iv) The conditional second moment $E(Z^2|X = x)$ is continuous on $[a, b]$. Moreover, $EZ^2 < \infty$.

Condition (i) ensures that the estimator (2.5) is well defined. Conditions (ii)–(iv) are analogous to those required in the ordinary nonparametric regression. Moreover, the rates depend on the following condition of the kernel function.

CONDITION 2. The kernel $K(\cdot)$ is a kth-order kernel. Namely,

$$\int_{-\infty}^{\infty} K(y)\,dy = 1, \qquad \int_{-\infty}^{\infty} y^k K(y)\,dy \neq 0,$$

$$\int_{-\infty}^{\infty} y^j K(y)\,dy = 0 \quad \text{for } j = 1, \dots, k-1.$$

Each of the following subsections contains two sets of results. The first set addresses the local and global rates of convergence. The second addresses the uniform results.

The global rates are described in terms of weighted L_p-norms which are defined as follows. Let $g(\cdot)$ denote a real-valued function on the line and let $w(\cdot)$ be a nonnegative weight function. Put

$$\|g(\cdot)\|_{wp} = \left\{\int |g(x)|^p w(x)\,dx\right\}^{1/p}, \qquad 1 \le p < \infty, \qquad \|g(\cdot)\|_{w\infty} = \sup_x |w(x)g(x)|.$$

To state the uniform results, we need to introduce a class of joint densities of (X, Z). In this class, Condition 1 should hold uniformly in order to obtain uniform results. More precisely, let B denote a positive constant and let $[a, b]$ be a compact interval. Denote the smallest integer exceeding $p/2$ by r_p: $r_p \ge p/2$. Define

$$\begin{aligned}\mathscr{F}_{k,B,p} = \Big\{ f(x,z) \colon |m^{(j)}(\cdot)| \le B,\ j = 1, \dots, k,\ |f_X^{(k)}(\cdot)| \le B, \\ E(|Z|^{r_p} | X = \cdot) \le B,\ \min_{a \le x \le b} f_X(x) \ge B^{-1} \Big\}.\end{aligned} \tag{3.3}$$

Note that this class $\mathscr{F}_{k,B,p}$ rephrases the standard conditions so that they hold uniformly. The condition $E(|Z|^{r_p}|X = \cdot) \le B$ is used only for the ordinary smooth case, and can be replaced by assuming bounded conditional density of X given Z.

3.1. *Super smooth error distributions.* The rates of convergence of kernel estimators under super smooth error models will be considered in this section. Let

$$b_k(x) \equiv (-1)^k \left[\frac{[m(x) f_X(x)]^{(k)}}{k!} - m(x) \frac{f_X^{(k)}(x)}{k!} \right] f_X^{-1}(x) \int_{-\infty}^{\infty} u^k K(u)\,du,$$

where f_X is the marginal density of X. The following result treats the local and global rates.

THEOREM 1. *Suppose that Conditions* 1 *and* 2 *hold and that the first half inequality of* (3.1) *is satisfied. Assume that* $\phi_K(t)$ *has a bounded support on* $|t| \le M_0$. *Then, for bandwidth* $h_n = c(\log n)^{-1/\beta}$ *with* $c > M_0(2/\gamma)^{1/\beta}$,

$$\begin{aligned} &E|(\hat{m}_n(x) - m(x)) \hat{f}_n(x)/f_X(x)|^2 \\ &\quad = (c^k b_k(x))^2 (\log n)^{-2k/\beta}(1 + o(1)) \end{aligned} \tag{3.4}$$

and

$$E\int_a^b \left|(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f_X(x)\right|^2 dx$$

$$= \int_a^b \left[c^k b_k(x)\right]^2 dx (\log n)^{-2k/\beta}(1 + o(1)).$$

The factor $\hat{f}_n(x)/f_X(x)(\to_P 1)$ is used to avoid the technical difficulty of the possibility of having 0 in the denominator of $\hat{m}(x) - m(x)$. It does not seriously affect the understanding of the statistical properties of the proposed estimator. Indeed, the proof of Theorem 1 also shows that

$$E\left((\hat{m}_n(x) - m(x))^2 | Y_1, \dots, Y_n\right) = \left(c^k b_k(x)\right)^2 (\log n)^{-2k/\beta}(1 + o_P(1)).$$

REMARK 1. Estimating regression functions in the presence of super smooth errors is extremely difficult, since the rates of convergence are very slow. Nevertheless, the variance of a kernel estimator can be very large (even going to infinity), if $c < M_0(2/\gamma)^{1/\beta}$, where c is the constant factor of the bandwidth. However, when $c > M_0(2/\gamma)^{1/\beta}$, the variance converges to 0 much faster than the bias does.

The preceding rates also hold uniformly over $\mathscr{F}_{k,B,2}$.

THEOREM 2. *Suppose that $\phi_\varepsilon(\cdot)$ and $K(\cdot)$ satisfy the conditions of Theorem 1. If the weight function $w(x)$ has a support $[a, b]$, then*

$$\lim_{d\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathscr{F}_{k,B,2}} P_f\left\{|\hat{m}_n(x) - m(x)| \ge d(\log n)^{-k/\beta}\right\} = 0$$

and

$$\lim_{d\to\infty} \limsup_{n\to\infty} \sup_{f\in\mathscr{F}_{k,B,2}} P_f\left\{\|\hat{m}_n(\cdot) - m(\cdot)\|_{wp} \ge d(\log n)^{-k/\beta}\right\} = 0, \qquad 1 \le p \le \infty.$$

An interesting feature of Theorem 2 is that $\hat{m}_n(\cdot)$ converges to $m(\cdot)$ with the same rates for both weighted L_p-loss $(1 \le p < \infty)$ and L_∞-loss. The result is not true for the ordinary nonparametric regression, where the global rates of convergence under L_∞-loss are slower [see Stone (1982)].

3.2. *Ordinary smooth error distributions.* This section considers kernel estimators under ordinary smooth error distributions. To compute the variance of the kernel density explicitly, we need the following condition on the tail of $\phi_\varepsilon(t)$, which is a variation of (3.2):

$$(3.5) \qquad t^\beta \phi_\varepsilon(t) \to c, \qquad |t^{\beta+1}\phi'_\varepsilon(t)| = O(1) \quad \text{as } t \to \infty,$$

for some constants $c \ne 0$.

THEOREM 3. *Suppose Conditions* 1 *and* 2 *hold and that*

$$\int_{-\infty}^{\infty} |t^{\beta+1}|(|\phi_K(t)| + |\phi_K'(t)|)\,dt < \infty, \qquad \int_{-\infty}^{\infty} |t^{\beta+1}\phi_K(t)|^2\,dt < \infty.$$

Then, under the ordinary smooth error distribution (3.5) *and* $h_n = dn^{-1/(2k+2\beta+1)}$ *with* $d > 0$,

$$E|(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f_X(x)|^2 = \left[b_k^2(x)h_n^{2k} + \frac{1}{nh_n^{1+2\beta}}v(x)\right](1 + o(1))$$
$$= O(n^{-2k/[2(k+\beta)+1]})$$

and

$$E\int_a^b |(\hat{m}_n(x) - m(x))\hat{f}_n(x)/f_X(x)|^2\,dx = O(n^{-2k/[2(k+\beta)+1]}),$$

where $v(x)$ *is defined by*

$$v(x) = \frac{1}{2\pi f_X^2(x)}\int_{-\infty}^{\infty}\left|\frac{t^\beta}{c}\right|^2 |\phi_K(t)|^2\,dt \int_{-\infty}^{\infty} \tau^2(x-v)\,f_X(x-v)\,dF_\varepsilon(v),$$

with $\tau^2(\cdot) = E((Z - m(x))^2|X = \cdot)$.

A reason for computing the bias and variance explicitly in Theorem 3 is that such a result will be useful for the bandwidth selection and the asymptotic normality of kernel regression estimators. To justify rate optimality, we need the preceding results to hold uniformly in a class of densities. Formally, we have the following theorem.

THEOREM 4. *Under the conditions of Theorem* 3 *on* $\phi_\varepsilon(\cdot)$, h_n *and* $K(\cdot)$, *if the weight function has a bounded support* $[a, b]$, *then*

$$\lim_{d\to\infty}\limsup_{n\to\infty}\sup_{f\in\mathscr{F}_{k,B,2}} P_f\{|\hat{m}_n(x) - m(x)| \ge dn^{-k/[2(k+\beta)+1]}\} = 0 \tag{3.6}$$

and

$$\lim_{d\to\infty}\limsup_{n\to\infty}\sup_{f\in\mathscr{F}_{k,B,p}} P_f\{\|\hat{m}_n(\cdot) - m(\cdot)\|_{wp} \ge dn^{-k/[2(k+\beta)+1]}\} = 0, \quad 1 \le p < \infty. \tag{3.7}$$

REMARK 2. For a regression function with a bounded kth derivative, Table 1 illustrates various rates (optimal local and global rates) of convergence according to the error distribution. The rate optimality will be justified in the next section. Note that the optimal rates are achieved by kernel estimators whose kernel and bandwidth satisfy the conditions of Theorems 2 and 4.

4. Rate optimality. It appears that the rates of convergence in the previous section are slower than the ordinary rates for nonparametric regression in the absence of errors. In particular, for super smooth error distribu-

TABLE 1

Error distribution	Rates of convergence	Error distribution	Rates of convergence
$N(0,1)$	$(\log n)^{-k/2}$	Gamma(α, p)	$n^{-k/(2k+2p+1)}$
Cauchy$(0,1)$	$(\log n)^{-k}$	Double exponential	$n^{-k/(2k+5)}$

tions such as the normal, the rates of the proposed estimators are extremely slow (see subsection 3.1). In this section we show that it is not possible to improve their performance, as far as rates of convergence are concerned. In other words, the rates of convergence presented in Section 3 are in fact an intrinsic part of regression problems with errors in variables, and are not an artifact of kernel estimators.

In order to justify the preceding claim, we need to make some restrictions on the distribution of the error variable ε. Note that the distribution function of ε is assumed known and the conditions we impose here can be easily checked. A formal statement of these conditions is given in Theorem 5, which deals with local and global lower rates for super smooth cases.

THEOREM 5. *Suppose that the characteristic function $\phi_\varepsilon(\cdot)$ of error variable ε satisfies the second half inequality of* (3.1) *and that*

$$P\{|\varepsilon - u| \le |u|^{\alpha_0}\} = O\big(|u|^{-(a-\alpha_0)}\big) \quad \text{as } u \to \infty, \tag{4.1}$$

for some $0 < \alpha_0 < 1$ and $a > 1 + \alpha_0$. Then, for any fixed point x, there exists a positive constant D_1 such that

$$\liminf_{n\to\infty} \inf_{T_n} \sup_{f\in\mathscr{F}_{k,B,2}} P_f\big\{|\hat{T}_n(x) - m(x)| > D_1(\log n)^{-k/\beta}\big\} > 0, \tag{4.2}$$

where $\inf_{\hat{T}_n}$ denotes the infimum over all possible estimators $\hat{T}_n$. Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{f\in\mathscr{F}_{k,B,2}} (\log n)^{2k/\beta} E_f|\hat{T}_n(x) - m(x)|^2 > 0. \tag{4.3}$$

Furthermore, if the weight function $w(\cdot)$ is positive and continuous on an interval, then

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{f\in\mathscr{F}_{k,B,2}} (\log n)^{k/\beta} E_f\|\hat{T}_n(\cdot) - m(\cdot)\|_{wp} > 0, \forall\, 1 \le p \le \infty. \tag{4.4}$$

The tail condition (4.1) is a technical condition used in the proof of the theorem. Note that this condition holds if $f_\varepsilon(y) = O(|y|^{-a})$, $a > 1$, where $f_\varepsilon(\cdot)$ is the density of ε and exists for all super smooth distributions. Theorem 5 includes the commonly used super smooth distributions such as normal, Cauchy and their mixtures as an error variable. The ordinary smooth cases, which include all gamma distributions and symmetric gamma distributions (e.g., double exponential distributions), are treated in the following theorem.

THEOREM 6. *Suppose that the characteristic function $\phi_\varepsilon(\cdot)$ of error variable ε satisfies*

$$\left|t^{-\beta-j}\phi_\varepsilon^{(j)}(t)\right| \le d \quad \text{for } j = 0, 1, 2.$$

Then, for any fixed point x, there exists a positive constant D_2 such that

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{f\in\mathscr{F}_{k,B,2}} P_f\left\{|\hat{T}_n(x) - m(x)| > D_2 n^{-k/[2(k+\beta)+1]}\right\} > 0. \tag{4.5}$$

Moreover,

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{f\in\mathscr{F}_{k,B,2}} n^{2k/[2(k+\beta)+1]} E_f|\hat{T}_n(x) - m(x)|^2 > 0. \tag{4.6}$$

Furthermore, if the weight function $w(\cdot)$ is positive and continuous on an interval, then

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{f\in\mathscr{F}_{k,B,p}} n^{k/[2(k+\beta)+1]} E_f\|\hat{T}_n(\cdot) - m(\cdot)\|_{wp} > 0, \forall\, 1 \le p < \infty. \tag{4.7}$$

REMARK 3. Statements (4.2) and (4.5) indicate that the rates obtained in Section 3 are indeed optimal.

The idea of establishing the lower bound is interesting and can be highlighted as follows. A detailed proof is deferred to Section 7. We use pointwise estimation (4.2) and (4.5) to illustrate the idea; the global lower bound can be treated similarly by combining the argument on pointwise estimation with the idea of the adaptively local one-dimensional subproblem of Fan (1993). Note that (4.3) and (4.6) follow directly from (4.2) and (4.5) via Chebyshev's inequality. The key of the proof is to reduce the lower bound for the regression problem to that of a density estimation problem so that the known results [Fan (1991a)] can be applied there.

To simplify our discussion, suppose that we wish to estimate the regression function at $x = 0$. Let f_0 and g_0 denote symmetric density functions such that [see (3.3)]

$$\min_{a\le x\le b} f_0(x) \ge B^{-1} \quad \text{and} \quad \int_{-\infty}^{\infty} |z|^{r_p} g_0(z)\,dz \le B. \tag{4.8}$$

Put $f_1(x,z) = f_0(x)g_0(z)$. Then $f_1 \in \mathscr{F}_{k,B,2}$. Now let $h_0(x)$ be a function satisfying

$$\int_{-\infty}^{\infty} z^j h_0(z)\,dz = j, \qquad j = 0, 1, \tag{4.9}$$

and let $\{a_n\}$ denote a sequence of positive numbers with $a_n \to 0$ [a_n is given by (4.13)]. Put

$$f_2(x,z) = f_0(x)g_0(z) + a_n^k H(x/a_n)h_0(z), \tag{4.10}$$

where the function $H(\cdot)$ will be specified in the proofs of Theorems 5 and 6 so that $f_2 \in \mathscr{F}_{k,B,2}$. By (4.8)–(4.10),

$$m_1(0) \equiv E_{f_1}(Z|X=0) = 0 \quad \text{and} \quad m_2(0) \equiv E_{f_2}(Z|X=0) = a_n^k H(0)/f_0(0).$$

Thus

$$\Delta \equiv |m_1(0) - m_2(0)|/2 = a_n^k H(0)/(2 f_0(0)). \tag{4.11}$$

Next, let

$$f_j * F_\varepsilon(y,z) = \int f_j(y-x,z)\, dF_\varepsilon(x),\ j = 1,2,$$

where F_ε is the cdf of the error distribution. If the χ^2-distance satisfies

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(f_1 * F_\varepsilon - f_2 * F_\varepsilon)^2}{f_1 * F_\varepsilon} dy\, dz = a_n^{2k} \int_{-\infty}^{\infty} \frac{h_0^2}{g_0} dz \left(\int_{-\infty}^{\infty} \frac{[H(\cdot/a_n) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy \right)$$
$$= O\left(\frac{1}{n}\right),$$

or, equivalently,

$$a_n^{2k} \int_{-\infty}^{\infty} \frac{[H(\cdot/a_n) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy = O\left(\frac{1}{n}\right), \tag{4.12}$$

then $\liminf_{n\to\infty} d_n > 0$ [see Donoho and Liu (1991)], where

$$d_n = \inf_{0 \le \hat{\phi} \le 1} \left(E_{f_1}\hat{\phi}_n + E_{f_2}(1 - \hat{\phi}_n) \right),$$

with ϕ being a test statistic based on the random sample. Therefore,

$$\inf_{\hat{T}_n} \sup_{f \in \mathscr{F}_{k,B,2}} P_f\{|\hat{T}_n(0) - m(0)| \ge \Delta\}$$
$$\ge \tfrac{1}{2} \inf_{\hat{T}_n} \left(P_{f_1}\{|\hat{T}_n(0) - m_1(0)| \ge \Delta\} + P_{f_2}\{|\hat{T}_n(0) - m_2(0)| \le \Delta\} \right)$$
$$\ge \tfrac{1}{2} d_n,$$

which is bounded away from 0. Hence a_n^k is the lower rate given in (4.3) or (4.5). It remains to determine a_n from (4.12).

If we use the same argument for estimating a density f_X at point 0 from the convolution model $Y = X + \varepsilon$, we will end up with the same problem of finding the largest (in rate) a_n from (4.12). The solution is found in Fan (1991a) and is given by

$$a_n = \begin{cases} \gamma^{-1/\beta}(\log n + c_1 \log\log n)^{-1/\beta}, & \text{for the super smooth case,} \\ c_2 n^{-1/(2k+2\beta+1)}, & \text{for the ordinary smooth case,} \end{cases} \tag{4.13}$$

where c_1 and c_2 are constants. This together with (4.11) gives the result on the lower rate.

One final remark: Our method of perturbation is quite different from those in the literature of nonparametric regression [see, e.g., Stone (1980, 1982) among others], where perturbation is applied directly to the regression function for some famous submodel (e.g., normal submodel). Our idea is to reduce explicitly the problem to a related density estimation problem so that some known facts from density estimation can be used. Indeed, the traditional construction *cannot* easily handle our more sophisticated errors-in-variables problems.

5. Simulations. In this section the sampling behavior of the deconvoluted kernel estimate (2.5) is examined by two simulated examples. In these examples X is a normal random variable and is observed through $Y = X + \varepsilon$, where the variance σ_0^2 of the error ε is chosen so that the reliability ratio [Fuller (1987)]:

$$r = \frac{\operatorname{Var}(X)}{\sigma_0^2 + \operatorname{Var}(X)} = 0.70.$$

Moreover, to study the effect of error distributions on the mean squared error of the estimator (2.5), ε is taken to be a normal and a double exponential random variable, respectively, in the following examples. Also, two different regression models are considered:

$$m_1(x) = x_+^3(1-x)_+^3 \quad \text{and} \quad m_2(x) = 1 + 4x.$$

EXAMPLE 1 (Normal errors). Let $(Y_1, Z_1), \dots, (Y_n, Z_n)$ denote a random sample from the distribution of (Y, Z), where

$$Y = X + \varepsilon, \qquad X \sim N(0.5, 0.25^2), \qquad \varepsilon \sim N(0, \sigma_0^2),$$

and Z is the response variable defined by

$$Z = m_1(X) + \epsilon, \qquad \epsilon \sim N(0, 0.0015^2)$$

or

$$Z = m_2(X) + \epsilon, \qquad \epsilon \sim N(0, 0.25^2).$$

In this case, $\phi_\varepsilon(t) = \exp(-\frac{1}{2}\sigma_0^2 t^2)$. Suppose the kernel function $K(\cdot)$ has a Fourier transform given by $\phi_K(t) = (1-t^2)_+^3$. By (2.4),

$$K_n(x) = \frac{1}{\pi}\int_0^1 \cos(tx)(1-t^2)^3 \exp\left(\frac{\sigma_0^2 t^2}{2h_n^2}\right) dt. \tag{5.1}$$

EXAMPLE 2 (Double exponential errors). We use the same model as in Example 1 except ε now has a double exponential distribution:

$$f_\varepsilon(z) = \left(\sigma_0\sqrt{2}\right)^{-1}\exp\left(-\sqrt{2}|z|/\sigma_0\right),$$

with the characteristic function

$$\phi_\varepsilon(t) = \frac{1}{1 + \frac{1}{2}\sigma_0^2 t^2}.$$

From (2.4),

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi_K(t)\left(1 + \frac{\sigma_0^2 t^2}{2h_n^2}\right) dt$$

$$= K(x) + \frac{\sigma_0^2 t^2}{2h_n^2}\frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)t^2\phi_K(t)\, dt$$

$$= K(x) - \frac{\sigma_0^2}{2h_n^2}K''(x).$$

If $K(\cdot)$ is further chosen to be the Gaussian kernel

$$K(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2),$$

then

$$K_n(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)\left[1 - \frac{\sigma_0^2}{2h_n^2}(x^2 - 1)\right]. \tag{5.2}$$

For each model, we use three different kernels: (5.1), (5.2) and the Gaussian (naive estimate by ignoring error). This is designed to test how much can be gained by using deconvolution and to examine the robustness of the deconvoluted kernel (e.g., normal deconvoluted kernel for double exponential error). Surprisingly, the deconvolution method is robust to the error assumption and is significantly better than the naive estimate, as illustrated in Tables 2 and 3. Note that as n increases, the ASE for the double exponential error decreases faster than that for the normal error. Moreover, the ASE for the double exponential error is smaller than the normal error. These are consistent with our theory.

For each simulation, we compute the average squared error (ASE) at 101 grid points from 0.1 to 0.9 using a geometric sequence of 21 bandwidths ranging in [0.1, 0.15] for $m_1(\cdot)$ and [0.02, 0.2] for $m_2(\cdot)$. The optimal bandwidth is selected to minimize the ASE among these 21 candidates. In a sense, this optimal bandwidth is the finite-sample optimal bandwidth. Notice that the preceding intervals are chosen wide enough so that the optimal bandwidth is

TABLE 2
SE ($\times 10^{-6}$) *for estimating the model*: $m(x) = x_+^3(1 - x)_+^3$

	Error distributions					
	Double exponential error			**Normal error**		
Kernel	$n = 200$	$n = 400$	$n = 800$	$n = 200$	$n = 400$	$n = 800$
(5.1)	4.0775	2.9173	2.3013	7.9870	6.7166	5.8625
(5.2)	4.4131	2.9248	2.5739	7.4348	7.0934	6.5009
Gaussian	8.1306	7.4412	7.1230	11.8280	11.2378	10.8662

TABLE 3
SE for estimating the model: $m(x) = 1 + 4x$

	Error distributions					
	Double exponential error			**Normal error**		
Kernel	$n = 200$	$n = 400$	$n = 800$	$n = 200$	$n = 400$	$n = 800$
(5.1)	0.04630	0.03075	0.02224	0.06292	0.05146	0.03920
(5.2)	0.05052	0.04222	0.02785	0.07004	0.05578	0.04573
Gaussian	0.08244	0.07165	0.06582	0.10339	0.09272	0.08717

contained in the interior of these intervals. Tables 2 and 3 report the average of these optimal ASE's in 100 replications.

6. Concluding remarks. To enhance the applicability of nonparametric regression, the present paper considers a new class of kernel estimators on which an examination of the effect of measurement errors is based. This new estimator is constructed by combining the ordinary kernel estimator and the idea of deconvolution in density estimation. The estimator is shown to possess various optimal properties that are characterized by the type of error distributions. Some insights have been gained in this investigation and are highlighted as follows:

1. The difficulty of nonparametric regression with errors in variables depends strongly on the smoothness of the error distribution: the smoother, the harder. This provides a new understanding of the intrinsic features of the problems in errors in variables.
2. As opposed to the approach to regression analysis with errors in variables based on normal error distributions, our study shows that this popular model suffers the drawback that the kernel estimators have extremely slow rates of convergence. We also show that this is an intrinsic part of the problem and is not an artifact of the kernel method.
3. For error distributions such as gamma or double exponential, the convergent rates of the modified kernel estimators are reasonable and behave very similarly to the usual kernel method. In fact, these results show that the usual kernel approach is a special case of our method.
4. Traditional arguments for establishing lower bounds for nonparametric regression estimators are difficult to generalize to the context of errors in variables. The current approach develops these bounds by reducing the regression problem to the corresponding density estimation problem via a new line of arguments.

7. Proofs. Let $f(x, z)$ and $g(y, z)$ denote the joint densities of (X, Z) and (Y, Z), respectively. By the independence of ε and (X, Z) and $Y = X + \varepsilon$,

$$g(y, z) = \int_{-\infty}^{\infty} f(y - x, z)\, dF_{\varepsilon}(x), \tag{7.1}$$

where $F_\varepsilon(\cdot)$ is the cdf of ε. We always denote the marginal density of X by $f_X(x)$.

7.1. *Proof of Theorem* 1. The proof of this theorem depends on the following lemma.

LEMMA 1. *If $\phi_K(\cdot)$ vanishes outside the interval $[-M_0, M_0]$, then*

$$EA_n(x) = \frac{1}{h_n}\int_{-\infty}^{\infty}[m(u) - m(x)]K\left(\frac{x-u}{h_n}\right)f_X(u)\,du,$$

where

$$A_n(x) = (nh_n)^{-1}\sum K_n\left(\frac{x-Y_j}{h_n}\right)[Z_j - m(x)]. \tag{7.2}$$

PROOF. According to (7.1) and (2.4),

$$\begin{aligned} h_nEA_n(x) &= EK_n((x-Y_1)/h_n)[Z_1 - m(x)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}K_n((x-y)/h_n)[z-m(x)]g(y,z)\,dy\,dz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2\pi}\exp(-it(x-y)/h_n)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} \\ &\qquad\times[z-m(x)]f(y-u,z)\,dt\,dF_\varepsilon(u)\,dy\,dz. \end{aligned}\tag{7.3}$$

Note that the Fourier transform of convolution is equal to the product of the transforms:

$$\begin{aligned} &\int_{-\infty}^{\infty}\exp(ity/h_n)\left(\int_{-\infty}^{\infty}f(y-u,z)\,dF_\varepsilon(u)\right)dy \\ &\qquad = \phi_\varepsilon(t/h_n)\int_{-\infty}^{\infty}\exp(ity/h_n)f(y,z)\,dy. \end{aligned}$$

Using this, we obtain

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-it(x-y)/h_n)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}f(y-u,z)\,dF_\varepsilon(u)\,dy\,dt \\ &\quad= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx/h_n)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} \\ &\qquad\times\left(\int_{-\infty}^{\infty}\exp(ity/h_n)\int_{-\infty}^{\infty}f(y-u,z)\,dF_\varepsilon(u)\,dy\right)dt \\ &\quad= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx/h_n)\left(\phi_K(t)\int_{-\infty}^{\infty}\exp(ity/h_n)f(y,z)\,dy\right)dt \\ &\quad= \int_{-\infty}^{\infty}K\left(\frac{x-y}{h_n}\right)f(y,z)\,dy, \end{aligned}\tag{7.4}$$

where the last equality follows from the inversion of the Fourier transform: The inversion of the products of two Fourier transforms equals the convolution. By (7.3) and (7.4),

$$EA_n(x) = \frac{1}{h_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[z - m(x)]K\left(\frac{x-y}{h_n}\right)f(y,z)\,dy\,dz$$
$$= \frac{1}{h_n}\int_{-\infty}^{\infty}[m(y) - m(x)]K\left(\frac{x-y}{h_n}\right)f_X(y)\,dy. \qquad \square$$

PROOF OF THEOREM 1. First of all,

$$(\hat{m}_n(x) - m(x))\frac{\hat{f}_n(x)}{f_X(x)} = \frac{A_n(x)}{f_X(x)}. \tag{7.5}$$

By Lemma 1, the "bias" is given by

$$\begin{aligned} EA_n(x) &= \frac{1}{h_n}\int_{-\infty}^{\infty}[m(u) - m(x)]K\left(\frac{x-u}{h_n}\right)f_X(u)\,du \\ &= f_X(x)b_k(x)h_n^k(1 + o(1)). \end{aligned} \tag{7.6}$$

Since X and ε are independent,

$$\begin{aligned} h_n^2\operatorname{Var}(A_n(x)) &= \frac{1}{n}\operatorname{Var}\left(K_n\left(\frac{x-Y}{h_n}\right)[Z - m(x)]\right) \\ &\le \frac{1}{n}E\left|K_n\left(\frac{x-Y}{h_n}\right)\right|^2[Z - m(x)]^2 \\ &= \frac{1}{n}EK_n^2\left(\frac{x-X-\varepsilon}{h_n}\right)\tau^2(X), \end{aligned} \tag{7.7}$$

where $\tau^2(X) = E((Z - m(x))^2|X)$. Next let us evaluate the first factor in (7.7). By the first half of (3.1), there exists a constant M such that

$$|\phi_\varepsilon(t)| > (d_0/2)|t|^{\beta_0}\exp(-|t|^\beta/\gamma) \quad \text{for } |t| > M.$$

Therefore, by the bounded support of $\phi_K(t)$,

$$\begin{aligned} \sup_x|K_n(x)| &\le \int_{-\infty}^{\infty}\frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}dt \\ &\le 2\int_0^{Mh_n}\frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}dt \\ &\quad + \frac{4}{d_0}\int_{Mh_n}^{M_0}|\phi_K(t)|\left|\frac{t}{h_n}\right|^{-\beta_0}\exp(|t/h_n|^\beta/\gamma)\,dt \\ &= O(h_n) + O\left(h_n^{-1}\exp(|M_0/h_n|^\beta/\gamma)\right). \end{aligned} \tag{7.8}$$

It follows from (7.7) and $h_n = c(\log n)^{-1/\beta}$ that

$$\text{(7.9)} \quad \operatorname{Var}(A_n(x)) = O\left(\frac{1}{nh_n^3}\exp\left(2|M_0/h_n|^{\beta}/\gamma\right)\right)E\tau^2(X) = o(h_n^{2k}).$$

Equation (3.4) now follows from the usual bias and variance decomposition. Since (7.6) and (7.9) hold uniformly in $x \in (a, b)$, the second conclusion is also valid. □

The proof of Theorem 2 depends on the follow lemma.

LEMMA 2. *Under the conditions of Theorem* 2,

$$\sup_{f \in \mathscr{F}_{k,B,2}} E\sup_x |\hat{f}_n(x) - f_X(x)|^p = o(1), \qquad p \geq 1,$$

where $\hat{f}_n$ *is defined by* (2.3).

PROOF. It follows from (2.2) that

$$E\hat{f}_n(x) = \int_{-\infty}^{\infty} f_X(u)\frac{1}{h_n}K\left(\frac{x-u}{h_n}\right)du.$$

Thus

$$\text{(7.10)} \qquad \sup_{\mathscr{F}_{k,B,2}} \sup_x |E\hat{f}_n(x) - f_X(x)| = o(1).$$

Next (2.2) leads to

$$\sup_x |\hat{f}_n(x) - E\hat{f}_n(x)| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}|\hat{\phi}_n(t) - \phi_Y(t)|dt.$$

It is easy to verify that there exists a constant c_p such that

$$E\left[|\hat{\phi}_n(t) - \phi_Y(t)|^p\right] \leq c_p n^{-p/2}.$$

Using this and Hölder's inequality, we have

$$\begin{aligned}
(2\pi)^p E\sup_x |\hat{f}_n(x) - E\hat{f}_n(x)|^p &\leq E\left(\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}|\hat{\phi}_n(t) - \phi_Y(t)|dt\right)^p \\
&\leq \left(\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}dt\right)^{p-1} \\
\text{(7.11)} \qquad &\quad \times\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}E\left[|\hat{\phi}_n(t) - \phi_Y(t)|^p\right]dt \\
&\leq c_p\left(\frac{1}{\sqrt{n}\,h_n}\int_{-\infty}^{\infty}\frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}dt\right)^p.
\end{aligned}$$

The conclusion follows from (7.8), (7.10) and (7.11). □

7.2. *Proof of Theorem* 2. By (7.5) and Lemma 2, we need to show

$$\sup_{f\in\mathscr{F}_{k,B,2}} E\sup_x|A_n(x)|^2 \le 2\sup_{f\in\mathscr{F}_{k,B,2}}\sup_x|EA_n(x)|^2 \tag{7.12}$$
$$+2\sup_{f\in\mathscr{F}_{k,B,2}} E\sup_x|A_n(x)-EA_n(x)|^2$$
$$=O\big((\log n)^{-2k/\beta}\big).$$

By (7.6),

$$\sup_{f\in\mathscr{F}_{k,B,2}}\sup_x|EA_n(x)|^2 = O\big(h_n^{2k}\big).$$

Thus we need only to argue that the second term in (7.12) has the right order. Set

$$U_j(t) = \exp(itY_j)Z_j - E\exp(itY_j)Z_j$$

and

$$V_j(t) = \exp(itY_j) - E\exp(itY_j).$$

Note that

$$|m(x)| \le \big(E(|Z|^{r_p}|X=x)\big)^{1/r_p} \le B^{1/r_p} \equiv B_1.$$

It follows from this together with (2.4) and (7.2) that

$$\Big(2\pi\sup_x|A_n(x)-EA_n(x)|\Big)^2$$
$$\le\left[\int_{-\infty}^{\infty}\left(\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|+\frac{B_1}{n}\left|\sum_{j=1}^n V_j(t)\right|\right)\left|\frac{\phi_K(th_n)}{\phi_\varepsilon(t)}\right|dt\right]^2.$$

By Hölder's inequality, the last display is bounded by

$$\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}dt\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}\left[\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|+\left|\frac{B_1}{n}\sum_{j=1}^n V_j(t)\right|\right]^2dt$$
$$\le 2\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}dt\int_{-\infty}^{\infty}\frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}\left[\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|^2\right. \tag{7.13}$$
$$\left.+\left|\frac{B_1}{n}\sum_{j=1}^n V_j(t)\right|^2\right]dt.$$

Note that

$$E\left|\frac{1}{n}\sum_{j=1}^n U_j(t)\right|^2 \le \frac{1}{n}EZ^2 \quad\text{and}\quad E\left|\frac{1}{n}\sum_{j=1}^n V_j(t)\right|^2 \le \frac{1}{n}.$$

By (7.13),

$$\sup_{f \in \mathscr{F}_{k,B,2}} E \sup_x |A_n(x) - EA_n(x)|^2 = O\left(\frac{1}{n}\left[\int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|} dt\right]^2\right).$$

By (7.8) and the choice of h_n, the last expression has order $o((\log n)^{-A})$ for any $A > 0$. This completes the proof of Theorem 2. □

7.3. *Proof of Theorem* 3. By (7.5), it suffices to compute the bias and the variance of $A_n(x)$ defined by (7.2). According to (7.6),

$$EA_n(x) = f_X(x) b_k(x) h_n^k (1 + o(1)). \tag{7.14}$$

The variance is given by

$$\operatorname{Var}(A_n(x)) = \frac{1}{n}\operatorname{Var}\left(h_n^{-1}K_n\left(\frac{x - Y}{h_n}\right)[Z - m(x)]\right)$$

$$= \frac{1}{n}E\left|h_n^{-1}K_n\left(\frac{x - Y}{h_n}\right)\right|^2 [Z - m(x)]^2 + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|h_n^{-1}K_n\left(\frac{x - u - v}{h_n}\right)\right|^2 \tau^2(u) f_X(u)\, dF_\varepsilon(v)\, du + o\left(\frac{1}{n}\right) \tag{7.15}$$

$$= \frac{1}{nh_n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_n^2(u)\tau^2(x - v - uh_n) f_X(x - v - uh_n)\, dF_\varepsilon(v)\, du + o\left(\frac{1}{n}\right).$$

Note that by (3.5) and the dominated convergence theorem,

$$h_n^\beta K_n(y) \to \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ity)\frac{t^\beta}{c}\phi_K(t)\, dt \overset{\text{def}}{=} J(y).$$

By Lemma 3 (to be given at the end of this section),

$$|h_n^\beta K_n(y)| \le \frac{C}{1 + |y|},$$

for some positive constant C. According to (7.15) and Lemma 2.1 of Fan (1991b),

$$\operatorname{Var}(A_n(x)) = \frac{1}{nh_n^{1+2\beta}}\int_{-\infty}^{\infty} J^2(u)\, du \int_{-\infty}^{\infty} \tau^2(x - v) f_X(x - v)\, dF_\varepsilon(v)[1 + o(1)].$$

By Parseval's identity,

$$\int_{-\infty}^{\infty} J^2(u)\, du = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{t^\beta}{c}\right|^2 |\phi_K(t)|^2\, dt.$$

Hence

$$\operatorname{Var}(A_n(x)) = \frac{1}{nh_n^{1+2\beta}} \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{t^\beta}{c} \right|^2 |\phi_K(t)|^2 \, dt$$
$$\times \int_{-\infty}^{\infty} \tau^2(x-v) f_X(x-v) \, dF_\varepsilon(v)[1 + o(1)].$$

The conclusion follows from bias and the variance decomposition.

LEMMA 3. *Under the conditions of Theorem* 3,

$$|h_n^\beta K_n(y)| \le \frac{C}{1+|y|},$$

for some constant C.

PROOF. According to (3.5), there are positive constants M and c_1 such that

$$|\phi_\varepsilon(t)| \ge c_1 |t|^{-\beta} \quad \text{for } |t| > M.$$

Therefore,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|} dt &\le 2\int_0^{Mh_n} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|} dt + 2\int_{Mh_n}^{\infty} \frac{|\phi_K(t)|}{c_1} \left| \frac{t}{h_n} \right|^\beta dt \\ &\le 2Mh_n \frac{\max|\phi_K(t)|}{\min|\phi_\varepsilon(t)|} + \frac{2h_n^{-\beta}}{c_1} \int_0^\infty |\phi_K(t)| t^\beta \, dt \\ &= O(h_n^{-\beta}). \end{aligned} \tag{7.16}$$

This together with (2.4) yields

$$|h_n^\beta K_n(y)| \le \frac{h_n^\beta}{2\pi} \int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|} dt = O(1). \tag{7.17}$$

Next, integration by parts and a similar argument in (7.16) lead to

$$|h_n^\beta K_n(y)| \le \frac{h_n^\beta}{2\pi |y|} \int_{-\infty}^{\infty} \left| \left(\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} \right)' \right| dt \le D|y|^{-1}, \tag{7.18}$$

where D is a positive constant. The desired conclusion follows from (7.17) and (7.18). □

7.4. *Proof of Theorem* 4. We start the proof with a lemma.

LEMMA 4. *Under the conditions of Theorem* 4,

$$\sup_{f \in \mathscr{F}_{k,B,2}} E \sup_x |\hat{f}_n(x) - f_X(x)|^p = o(1).$$

PROOF. This follows easily from (7.10), (7.11) and (7.16). □

PROOF OF THEOREM 4. The local rate (3.6) follows the similar argument given in Theorem 3. We focus on proving the global results (3.7). By Lemma 4 and (7.2), it suffices to show

$$\sup_{f \in \mathscr{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)\,dx = O\big(n^{-pk/[2(k+\beta)+1]}\big).$$

By Lemma 1 and (7.14),

$$\sup_{f \in \mathscr{F}_{k,B,p}} \sup_x EA_n(x)| = O\big(h_n^k\big).$$

Thus

$$(7.19)\qquad \begin{aligned} &\sup_{f \in \mathscr{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)\,dx \\ &\quad \le 2^p \sup_{f \in \mathscr{F}_{k,B,p}} E\int_a^b |A_n(x) - EA_n(x)|^p w(x)\,dx \\ &\quad\quad + O\big(n^{-pk/[2(k+\beta)+1]}\big). \end{aligned}$$

Hence we need only to justify that the first term of (7.19) is of the right order $O(n^{-pk/[2(k+\beta)+1]})$.

Recall that $r \equiv r_p$ = the smallest integer exceeding $p/2$ and put

$$\begin{aligned} T_j(x) \equiv T_{n,j}(x) &= h_n^{-1}K_n\big((x - Y_j)/h_n\big)\big[Z_j - m(x)\big] \\ &\quad - h_n^{-1}EK_n\big((x - Y_j)/h_n\big)\big[Z_j - m(x)\big]. \end{aligned}$$

Then

$$(7.20)\qquad E|A_n(x) - EA_n(x)|^p = E\left|\frac{1}{n}\sum_{j=1}^n T_j(x)\right|^p \le \left(E\left|\frac{1}{n}\sum_{j=1}^n T_j(x)\right|^{2r}\right)^{p/2r}.$$

Moreover,

$$(7.21)\qquad \sup_{f \in \mathscr{F}_{k,B,p}} \sup_{x \in [a,b]} E\left(\frac{1}{n}\sum_{j=1}^n T_j(x)\right)^{2r} = O\left(\left(\frac{1}{n}h_n^{1-2(1+\beta)}\right)^r\right).$$

[The proof of (7.21) will be given shortly.] The conclusion of the theorem follows from (7.19)–(7.21). □

We now prove (7.21) by a pair of lemmas, which hold uniformly in $f \in \mathscr{F}_{k,B,p}$.

LEMMA 5. *Under the conditions of Theorem* 4,

$$\sup_x E|T_1(x)|^l = O\big(h_n^{1-l(\beta+1)}\big) \quad \text{for } l = 2,\dots,r.$$

PROOF. Let $\nu_l(x) = E(|Z|^l|X = x)$. Then $\nu_l(x) \le B$, by (3.3). It follows from the inequality $|a + b|^l \le 2^l(|a|^l + |b|^l)$ that

$$
\begin{aligned}
h_n^l E|T_1(x)|^l &\le 2^{l+1}E|K_n((x - Y)/h_n)[Z - m(x)]|^l \\
&= 2^{2l+1}\big[E|K_n((x - Y)/h_n)|^l \nu_l(X) \\
&\qquad + B_1^l E|K_n((x - Y)/h_n)|^l\big] \\
&\le 2^{2l+1}(B + B_1^l)E|K_n((x - Y)/h_n)|^l.
\end{aligned}
\tag{7.22}
$$

Recall that $f_Y(y)$ is the density of $Y = X + \varepsilon$. Then, by Lemma 4,

$$
\begin{aligned}
E|K_n((x - Y)/h_n)|^l &= h_n \int_{-\infty}^{\infty} |K_n(y)|^l f_Y(x - yh_n)\,dy \\
&\le C^l h_n^{1-l\beta} \int_{-\infty}^{\infty} \frac{1}{(1 + |y|)^l} f_Y(x - yh_n)\,dy \\
&= O(h_n^{1-l\beta}).
\end{aligned}
\tag{7.23}
$$

The desired result follows from (7.22) and (7.23). □

LEMMA 6. *Under the conditions of Theorem* 4,

$$\sup_{x\in[a,b]} E\left(\frac{1}{n}\sum_{j=1}^{n} T_j(x)\right)^{2r} = O\left(\left(\frac{1}{n}h_n^{1-2(1+\beta)}\right)^r\right).$$

PROOF. Write $T_j = T_j(x)$. By the multinominal formula,

$$\left(\sum_{j=1}^{n} T_j(x)\right)^{2r} = \sum_{k=1}^{2r} \sum{}' \frac{(2r)!}{r_1! \cdots r_k!} \frac{1}{k!} \sum{}'' T_{j_1}^{r_1} \cdots T_{j_k}^{r_k},$$

where Σ' sums over k-tuples of positive integers $(r_1, \dots, r_k)$ satisfying $r_1 + \cdots + r_k = 2r$ and Σ'' extends over k-tuples of distinct integers $(j_1, \dots, j_k)$ in the range $1 \le j \le n$.

By independence and that T_j has mean 0,

$$E\left(\sum_{j=1}^{n} T_j(x)\right)^{2r} = \sum_{k=1}^{2r} \sum{}''' \frac{(2r)!}{r_1! \cdots r_k!} \frac{1}{k!} \sum{}'' E(T_{j_1}^{r_1}) \cdots E(T_{j_k}^{r_k}),$$

where Σ''' sums over k-tuples of positive integers $(r_1, \dots, r_k)$ satisfying $r_1 + \cdots + r_k = 2r$ and $r_j \ge 2$, $j = 1, \dots, k$. Thus $k \le r$. By Lemma 5,

$$
\begin{aligned}
\sum{}'' E(T_{j_1}^{r_1}) \cdots E(T_{j_k}^{r_k}) &\le n^k E(T_{j_1}^{r_1}) \cdots E(T_{j_k}^{r_k}) \\
&\le n^k O(h_n^{1-r_1(\beta+1)}) \cdots O(h_n^{1-r_k(\beta+1)}) \\
&= O\left(n^r (h_n^{1-2(\beta+1)})^r \frac{1}{(nh_n)^{r-k}}\right) \\
&= O\left(n^r (h_n^{1-2(\beta+1)})^r\right),
\end{aligned}
$$

since $nh_n \to \infty$. The desired result follows. This completes the proof of Lemma 6. □

7.5. *Proofs of Theorems* 5 *and* 6. We first justify the local lower rates of Theorems 5 and 6, that is, (4.2) and (4.5). The basic idea is outlined in Section 4. For simplicity, we prove only the case $x = 0$ in (4.2) and (4.5).

We now specify the functions f_0, g_0, h_0 and H in the heuristic argument of Section 4. Define

$$f_0(x) = C_r(1 + x^2)^{-r}, \qquad r > 0.5, \qquad g_0(z) = (\sqrt{2\pi}\, b)^{-1} \exp\left(-\frac{z^2}{2b^2}\right) \tag{7.24}$$

and

$$h_0(z) = \frac{1}{\sqrt{2\pi}}\left(\exp\left(-(z-1)^2/2\right) - \exp(-z^2/2)\right), \tag{7.25}$$

where $C_r = \int_{-\infty}^{\infty}(1 + x^2)^{-r}\,dx$, and b and r will be chosen later. Note that h_0 satisfies (4.9). We now construct the function $H(\cdot)$ as follows. Take a nonnegative symmetric function $\phi(t)$, which vanishes when $|t| \notin [1, 2]$ and has continuous m_0th derivatives, for some given m_0. Moreover,

$$\int_1^2 \phi(t)\,dt \neq 0.$$

Let $H(\cdot)$ be the Fourier inversion of $\phi(t)$:

$$H(x) = \frac{1}{\pi}\int_1^2 \cos(tx)\phi(t)\,dt.$$

Then $H(\cdot)$ has the following properties:

1. $H(0) \neq 0$.
2. $H(x)$ has all bounded derivatives.
3. $|H(x)| \le c_0(1 + x^2)^{-m_0/2}$, for some constant $c_0 > 0$.
4. $\phi_H(t) = 0$, when $|t| \notin [1, 2]$, where ϕ_H is the Fourier transform of H.

By the proper choice of r and b, the pair of densities f_1 and f_2 defined by (4.10) will be members of $\mathscr{F}_{k, B, 2}$. By the argument in Section 4, a_n^k would be the lower rates if a_n satisfies (4.12). According to Fan (1991a) and the conditions of Theorem 5, there is a positive constant c_1 such that the solution of a_n to (4.12) is given by

$$a_n = \left(\log n + c_1 \log(\log n)\right)^{-1/\beta}\gamma^{-1/\beta}. \tag{7.26}$$

Similarly, for Theorem 6, the solution is given by

$$a_n = c_2 n^{-1/(2k+2\beta+1)}, \tag{7.27}$$

where c_2 is a positive constant. Thus the conclusion of (4.2) and (4.6) follows.

Now we turn to the global rates (4.4) and (4.7). We use the idea of the adaptively local one-dimensional subproblem of Fan (1993). Specifically, see Theorem 1 of that paper. In the following discussion, we may assume that

$w(x) > 0$ on $[0,1]$. Let m_n denote a sequence of positive integers tending to ∞ and let $x_j = j/m_n$, $j = 1, 2, \dots, m_n$, be a grid point of $[0,1]$. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{m_n}) \in \{0,1\}^{m_n}$ be a vector of 0 and 1. Construct a sequence of functions:

$$m_{\boldsymbol{\theta}}(x) = m_n^{-k} \sum_{j=1}^{m_n} \theta_j H(m_n(x - x_j)).$$

Define a family of densities:

$$f_{\boldsymbol{\theta}}(x, z) = f_0(x) g_0(z) + \delta m_{\boldsymbol{\theta}}(x) h_0(z),$$

where f_0, g_0 and h_0 are defined by (7.24) and (7.25).

For suitable choice of $b > 0$, $r > 0.5$ and $\delta > 0$, we now show that $f_{\boldsymbol{\theta}}(x, z) \in \mathscr{F}_{k,B,p}$, which is a subset of $\mathscr{F}_{k,B,2}$. It is easy to see that $|h_0(z)| \le c_3 g_0(z)$ for all z, and, by Lemma 7 (to be given shortly), $|m_{\boldsymbol{\theta}}(x)| \le c_4(1 + x^2)^{-m_0/2}$. Thus, for sufficiently small δ and r, $f_{\boldsymbol{\theta}}$ is a density function satisfying

$$f_{\boldsymbol{\theta}}(x, z) \ge 0.5 f_0(x) g_0(z) \qquad \forall\, \boldsymbol{\theta} \in \{0,1\}^{m_n}. \tag{7.28}$$

Now the conditional mean is given by

$$E_{f_{\boldsymbol{\theta}}}(Z|X = x) = \delta m_{\boldsymbol{\theta}}(x)/f_0(x).$$

By Lemma 7 again, the kth derivative of the conditional expectation is bounded by the constant B for small $\delta > 0$. Similarly, other conditions in (3.3) are satisfied with a suitable choice of r and δ. Hence $f_{\boldsymbol{\theta}}(x, z) \in \mathscr{F}_{k,B,p}$ for all $\boldsymbol{\theta} \in \{0,1\}^{m_n}$.

Denote

$$\boldsymbol{\theta}_{j0} = (\theta_1, \dots, \theta_{j-1}, 0, \theta_{j+1}, \dots, \theta_{m_n})$$

and

$$\boldsymbol{\theta}_{j1} = (\theta_1, \dots, \theta_{j-1}, 1, \theta_{j+1}, \dots, \theta_{m_n}).$$

Then there is a positive constant c_5 so that the difference

$$\begin{aligned}\left|E_{f_{\boldsymbol{\theta}_{j0}}}(Z|X = x) - E_{f_{\boldsymbol{\theta}_{j1}}}(Z|X = x)\right| &= \delta m_n^{-k} |H(m_n(x - x_j))|/f_0(x) \\ &\ge c_5 |H(m_n(x - x_j))| m_n^{-k} \quad \text{for } x \in [0,1].\end{aligned}$$

Put

$$f_{\boldsymbol{\theta}_{j0}} * F_\varepsilon(y, z) = \int_{-\infty}^{\infty} f_{\boldsymbol{\theta}_{j0}}(y - x, z)\, dF_\varepsilon(x),$$

$$f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon(y, z) = \int_{-\infty}^{\infty} f_{\boldsymbol{\theta}_{j1}}(y - x, z)\, dF_\varepsilon(x).$$

By Theorem 1 of Fan (1993), if

$$\max_{1 \le j \le m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2\left(f_{\boldsymbol{\theta}_{j0}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon\right) \le c_6/n, \tag{7.29}$$

then

$$\inf_{\hat{T}_n(x)} \sup_{f \in \mathscr{F}_{k,B,p}} E_f \int_0^1 |\hat{T}_n(x) - m(x)|^p w(x)\, dx$$

$$\geq \frac{1 - \sqrt{1 - \exp(-c_6)}}{2^{p+1}} \int_0^1 w(x)\, dx \int_0^1 |H(x)|^p\, dx \left(c_5 m_n^{-k}\right)^p.$$

Thus m_n^{-k} is the global lower rate.

Now we determine m_n from (7.29). By (7.28), we have

$$\max_{1 \leq j \leq m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2\left(f_{\theta_{j0}} * F_\varepsilon, f_{\theta_{j1}} * F_\varepsilon\right)$$

$$\leq 2 m_n^{-2k} \int_{-\infty}^{\infty} h_0^2(z)/g_0(z)\, dz \int_{-\infty}^{\infty} \frac{\left[H(m_n(\cdot - x_j)) * F_\varepsilon\right]^2}{f_0 * F_\varepsilon} dy. \tag{7.30}$$

Note that there exists a positive constant c_7 such that $f_0(x) > c_7 f_0(x - x_j)$ for all $x_j \in [0, 1]$. Using this fact in (7.30) with a change of variable, we have

$$\max_{1 \leq j \leq m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2\left(f_{\theta_{j0}} * F_\varepsilon, f_{\theta_{j1}} * F_\varepsilon\right)$$

$$\leq 2 c_7^{-1} m_n^{-2k} \int_{-\infty}^{\infty} h_0^2(z)/g_0(z)\, dz \int_{-\infty}^{\infty} \frac{\left[H(m_n(\cdot - x_j)) * F_\varepsilon\right]^2}{f_0(\cdot - x_j) * F_\varepsilon} dx$$

$$= 2 c_7^{-1} m_n^{-2k} \int_{-\infty}^{\infty} h_0^2(z)/g_0(z)\, dz \int_{-\infty}^{\infty} \frac{\left[H(m_n(\cdot)) * F_\varepsilon\right]^2}{f_0 * F_\varepsilon} dx.$$

In other words, we need to determine m_n from

$$m_n^{-2k} \int_{-\infty}^{\infty} \frac{\left[H(m_n(\cdot)) * F_\varepsilon\right]^2}{f_0 * F_\varepsilon} dx = O\left(\frac{1}{n}\right). \tag{7.31}$$

Problem (7.31) is exactly the same as problem (4.12), by thinking of $a_n = m_n^{-1}$. The conclusion follows again from Fan (1991a) [see also (7.26) and (7.27)]. □

LEMMA 7. *Suppose that the function $G(x)$ satisfies*

$$|G(x)| \leq C(1 + x^2)^{-m}, \qquad m > 0.5.$$

Then there exists a positive constant C_1 such that, for any sequence $m_n \to \infty$,

$$\sum_{j=1}^{m_n} |G(m_n x - j)| \leq C_1 (1 + x^2)^{-m}.$$

PROOF. If $|x| \geq 2$, then $|m_n x - j| \geq m_n(|x| - 1)$ and

$$\sum_{j=1}^{m_n} |G(m_n x - j)| \leq C \sum_{j=1}^{m_n} \left(1 + m_n^2(|x| - 1)^2\right)^{-m}$$

$$= C m_n^{1-2m}(|x| - 1)^{-2m}$$

$$\leq C_1(1 + x^2)^{-m},$$

for some constant C_1. When $|x| < 2$,

$$\sum_{j=1}^{m_n} |G(m_n x - j)| \leq C \sum_{j=1}^{m_n} \left(1 + (m_n x - j)^2\right)^{-m} = O(1),$$

as was to be shown. □

Acknowledgments. We wish to thank Chuck Stone, an Associate Editor and the referees, who each made several helpful suggestions that improved the presentation and the results of the paper, in particular, the simulation study in Section 5.

REFERENCES

ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

BICKEL, P. J. and RITOV, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.* **15** 513–540.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

CARROLL, R. J., SPIEGELMAN, C. H., LAN, K. K. G., BAILEY, K. T. and ABBOTT, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* **71** 19–25.

DONOHO, D. L. and LIU, R. G. (1991). Geometrizing rates of convergence II and III. *Ann. Statist.* **19** 633–701.

FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1991b). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** 97–110.

FAN, J. (1993). Adaptively local one-dimensional subproblem with applications to a deconvolution problem. *Ann. Statist.* **21** 600–610.

FULLER, W. A. (1987). *Measurement Error Models.* Wiley, New York.

LIU, M. C. and TAYLOR, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17** 427–438.

PRENTICE, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *J. Amer. Statist. Assoc.* **81** 321–327.

SCHAFER, D. W. (1987). Covariate measurement error in generalized linear models. *Biometrika* **74** 385–391.

STEFANSKI, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72** 583–592.

STEFANSKI, L. A. and CARROLL, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13** 1335–1351.

STEFANSKI, L. A. and CARROLL, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74** 703–716.

STEFANSKI, L. A. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

STEFANSKI, L. A. and CARROLL, R. J. (1991). Deconvolution-based score tests in measurement error models. *Ann. Statist.* **19** 249–259.

STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

WHITTEMORE, A. S. (1989). Errors-in-variables regression using Stein estimates. *Amer. Statist.* **43** 226–228.

WHITTEMORE, A. S. and KELLER, J. B. (1988). Approximations for regression with covariate measurement error. *J. Amer. Statist. Assoc.* **83** 1057–1066.

ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions *Ann. Statist.* **18** 806–831.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27899-3260

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-7400