NONPARAMETRIC REGRESSION WITH ERRORS IN VARIABLES

Jianqing Fan
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27514

Young K. Truong
Department of Biostatistics
University of North Carolina
Chapel Hill, N.C. 27514

March 25, 1990

Abstract

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 covariates, which includes ordinary smooth regression functions as well as covariates with distributions satisfying regularity conditions. Furthermore, to achieve optimality, we show that the convergence rates of all nonparametric estimators have a lower bound possessed by the kernel estimators.

[0] *Abbreviated title.* Error-in-variable regression

AMS 1980 subject classification. Primary 62G20. Secondary 62G05, 62J99.

Key words and phrases. Nonparametric regression; Kernel estimator; Errors in variables; Optimal rates of convergence; Deconvolution.

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 et al (1984), Stefanski and Carroll (1985), Stefanski (1985), Fuller (1986), Prentice (1986), Bickel and Ritov (1987), Stefanski and Carroll (1987a), Schafer (1987), Whittemore and Keller (1988), and Whittemore (1989). However, the latter has centered around the parametric approach. That is, the regression function is assumed to take on a particular functional form. Attempts to unify these two approaches form the subject of this paper.

Let (X, Z) denote a pair of random variables and suppose it is desired to estimate the regression function $m(x) = E(Z|X = x)$. However, due to the measuring mechanism or the nature of the environment, the variable X is measured with error and is not directly observable [Fuller (1986), p.2]. Instead, $Y = X + \varepsilon$ is observed, where ε is a random disturbance whose distribution is known, and is independent of (X, 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) How well does it behave? (c) Does it possess some optimalities? The discussions of the issues center the core of the paper.

Suppose that $(Y_1, Z_1), \ldots, (Y_n, Z_n)$ is a random sample from the distribution of (Y, Z). We address the first issue by considering the following kernel type estimator

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

where $W_{n,j}(Y_1, \ldots, Y_n)$ is a weight function normalized so that $\sum_j W_{n,j}(Y_1, \ldots, Y_n) = 1$. These weights are constructed so that they will account for the errors in the covariate X. The idea is related with that of density estimation using deconvolution techniques. See Stefanski and Carroll (1987b), Fan (1988) and Section 2 for more details.

The second issue is addressed through two types of error distributions. An error is called *ordinary smooth* if the tails of its characteristic function decay to zero at an algebraic rate. It is called *super smooth* if its characteristic function has tails approaching zero exponentially fast. See Fan (1988) and Section 3 for a formal definition. For example, distributions such as double exponential and gamma are ordinary smooth, while normal and Cauchy are super smooth. The current paper examines to what extend that the distribution of ε affects the rates of convergence of the above nonparametric estimators, both locally and globally.

Depending whether the error is ordinary smooth or super smooth, the rates of convergence of the kernel estimators are quite different—the local and global rates are slower in the super smooth model while they are faster in the ordinary smooth model. These results also hold uniformly over a class of joint distributions of (X, Z) which includes regression functions possessing smoothness conditions, and covariates with distributions satisfying regularity conditions. For more details, see Section 3.

An interesting consequence of the results in Section 3 is worth mentioning. Our error-in-variable model includes the usual nonparametric regression model in the absence of error. Thus, as a corollary of Theorem 4, we show that kernel estimators attain optimal global rates of convergence under weighted L_p-loss ($1 \leq p < \infty$). The result, to our knowledge, appears to be new even in the ordinary nonparametric regression.

The third issue is focused on rate optimality. We construct minimax lower bounds on the rates of convergence—both locally and globally. The dependence of the lower rates on the smoothness of error distribution is clearly addressed.

The rates of convergence of the kernel estimators can also be characterized through the error distribution. Indeed, in Section 4, we will show that these rates provide lower bounds for all nonparametric regression function 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).

In contrast with previous results in parametric regression involving errors in variables, our investigation shows that one should be cautious about using normal as an error distribution, since the optimal estimators based on normal errors have very slow rates of convergence.

The paper is outlined as follows. Section 2 describes the idea of deconvolution and kernel estimators. Our assumptions and issues on rates of convergence are presented in Section 3. Section 4 deals with optimality. Section 5 contains further remarks. Proofs are given in Section 6.

2 Kernel estimators

Let $(X_1, Z_1), \ldots, (X_n, Z_n)$ 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 $K((x - X_j)/h_n)$:

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

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

Since the variables $X_1, \ldots, X_n$ are *not observable*, the kernel estimator $\hat{f}_n(x)$ will be constructed from $Y_j = X_j + \varepsilon_j$, $j = 1, \ldots, n$. 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

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(y - x) dF_\varepsilon(x). \tag{2.2}$$

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 (1987b) and Fan (1988) consider the following estimator:

$$\hat{f}_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_K(th_n) \frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)}\, dt, \tag{2.3}$$

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:

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_1^n \exp(itY_j). \tag{2.4}$$

Note that this estimator can be rewritten in the kernel form

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_1^n K_n(\frac{x - Y_j}{h_n}), \tag{2.5}$$

with

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\, dt. \tag{2.6}$$

Some other recent contributions to nonparametric deconvolution include Carroll and Hall (1988), Fan (1990), Liu and Taylor (1990), Zhang (1989), etc.

Eqs (2.1), (2.5) and (2.6) motivate the following regression function estimator involving errors in variables:

$$\begin{aligned} \hat{m}_n(x) &= \sum_j K_n(\frac{x - Y_j}{h_n})Z_j / \sum_i K_n(\frac{x - Y_i}{h_n}) \\ &= \frac{1}{nh_n}\sum_j K_n(\frac{x - Y_j}{h_n})Z_j / \hat{f}_n(x), \end{aligned} \tag{2.7}$$

where $\hat{f}_n(x)$ is defined by (2.5). Note that it is a kernel type estimator with kernel weight proportional to $K_n((x - Y_j)/h_n)$.

3 Performance of kernel estimators

The sampling behaviors of the kernel estimators (2.7) 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:

- Super smooth of order β: if the characteristic function of the error distribution $\phi_\varepsilon(\cdot)$ satisfies

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

  where d_0, d_1, β, γ are positive constants and β_0, β_1 are constants.

- 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, β.

For example,

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

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

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

Condition 1.

i. Let $a < b$. The marginal density $f_X(\cdot)$ of the unobserved X is bounded away from zero on the interval $[a, b]$, and has a bounded kth derivative.

ii. The characteristic function of X is absolutely integrable.

iii. The characteristic function of the error distribution $\phi_\varepsilon(\cdot)$ does not vanish.

iv. The regression function $m(\cdot)$ has a bounded kth derivative.

v. The conditional variance of $\sigma^2(x) = \mathrm{Var}(Z|X = x)$ is bounded and continuous.

The rates depend on the following condition of the kernel function:

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

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

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

Each of the following subsections contains two sets of results. The first set discusses 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 $\mathbf{R}$ and let $w(\cdot)$ be a non-negative weight function. Put

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

To state the uniform results, we need to introduce a class of joint densities of (X, Z). Let B denote a positive constant and $[a, b]$ be a compact interval. Denote the smallest integer exceeding $p/2$ by r_p: $r_p \ge p/2$. Define

$$\mathcal{F}_{k,B,p} = \{f(x,z) \;:\; |m^{(k)}(\cdot)| \le B, |m(\cdot)| \le B, |f_X^{(k)}(\cdot)| \le B, \int_{-\infty}^{\infty} |\phi_X(t)|\,dt \le B,$$
$$E(|Z|^{r_p}|X = x) \le B, \min_{a\le x\le b} f_X(x) \ge B^{-1}\}. \tag{3.3}$$

3.1 Super smooth error distributions

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 – 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 $c > M_0(2/\gamma)^{1/\beta}$,*

$$E|\hat{m}_n(x) - m(x)|^2 = (c^k b_k(x))^2 (\log n)^{-2k/\beta}(1 + o(1)), \tag{3.4}$$

and

$$E\int_a^b |\hat{m}_n(x) - m(x)|^2\,dx = \int_a^b \left[c^k b_k(x)\right]^2 dx (\log n)^{-2k/\beta}(1 + o(1)). \tag{3.5}$$

Remark 1. Estimating regression functions in the presence of super smooth errors are extremely difficult, since the rates of convergence are very slow. Nevertheless, the variance of a kernel estimator can be very large (even goes to infinite), 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 above result also holds uniformly over $\mathcal{F}_{k,B,2}$:

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*

$$\sup_{f\in\mathcal{F}_{k,B,2}} E|\hat{m}_n(x) - m(x)|^2 = O((\log n)^{-2k/\beta}), \tag{3.6}$$

and

$$\sup_{f\in\mathcal{F}_{k,B,2}} E\|\hat{m}_n(\cdot) - m(\cdot)\|_{wp} = O((\log n)^{-k/\beta}), \qquad 1 \le p \le \infty. \tag{3.7}$$

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 $\phi_\varepsilon(t)$:

$$t^\beta \phi_\varepsilon(t) \to c, \qquad |t^{\beta+1}\phi'_\varepsilon(t)| = O(1), \qquad \text{as } t \to \infty \tag{3.8}$$

for some constants $c \neq 0$.

Theorem 3. *Suppose Conditions 1 - 2 hold and that*

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

Then, under the ordinary smooth error distribution (3.8) and $h_n = dn^{-1/(2k+2\beta+1)}$ *with* $d > 0$,

$$
\begin{aligned}
E|\hat{m}_n(x) - m(x)|^2 &= \left[b_k^2(x)h_n^{2k} + \frac{1}{nh_n^{1+2\beta}}v(x)\right](1 + o(1)) \\
&= O\left(n^{-\frac{2k}{2(k+\beta)+1}}\right), \qquad (3.9)
\end{aligned}
$$

and

$$
E\int_a^b |\hat{m}_n(x) - m(x)|^2\,dx = O\left(n^{-\frac{2k}{2(k+\beta)+1}}\right), \qquad (3.10)
$$

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

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

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

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

$$
\sup_{f\in\mathcal{F}_{k,B,2}} E|\hat{m}_n(x) - m(x)|^2 = O\left(n^{-\frac{2k}{2(k+\beta)+1}}\right), \qquad (3.11)
$$

and

$$
\sup_{f\in\mathcal{F}_{k,B,p}} E\|\hat{m}_n(x) - m(x)\|_{wp}^p = O\left(n^{-\frac{kp}{2(k+\beta)+1}}\right), \qquad 1 \le p < \infty. \qquad (3.12)
$$

A direct consequence of (3.12) yields

$$
\sup_{f\in\mathcal{F}_{k,B,p}} E\|\hat{m}_n(x) - m(x)\|_{wp} = O\left(n^{-\frac{k}{2(k+\beta)+1}}\right), \qquad 1 \le p < \infty.
$$

Remark 2. For a regression function with a bounded k-th derivative, the following table illustrates various rates (optimal local and global rates) of convergence according to the error distribution. The rate optimality will be justified in next section.

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	$n^{-k/(2k+5)}$

Note that the optimal rates are achieved by kernel estimators whose kernel and bandwidth satisfy the conditions of Theorem 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 distributions such as the normal, the rates of the proposed estimators are extremely slow (see Section 3.1). In this section, we show that it is not possible to improve their performances, 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 claim above, 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 (see examples and Remark 2 in Section 3). A formal statement of these conditions is given in Theorem 5 below, 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*

$$P\left\{|\varepsilon - u| \le |u|^{\alpha_0}\right\} = O\left(|u|^{-(a-\alpha_0)}\right), \text{ as } u \to \infty$$

for some $0 < \alpha_0 < 1$ and $a > 1.5 + \alpha_0$. Then, for any fixed point x, there exists a positive constant D_1 such that

$$\inf_{\hat{T}_n} \sup_{f \in \mathcal{F}_{k,B,2}} E_f|\hat{T}_n(x) - m(x)|^2 > D_1 (\log n)^{-2k/\beta}, \tag{4.1}$$

here $\inf_{\hat{T}_n}$ *denotes the infimum over all possible estimators* $\hat{T}_n$. *Moreover, if the weight function* $w(\cdot)$ *is positive and continuous on an interval, then there exists a positive constant* D_2 *such that*

$$\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_{k,B,2}} E_f\|\hat{T}_n(\cdot) - m(\cdot)\|_{wp} > D_2(\log n)^{-k/\beta}, \quad \forall 1 \le p \le \infty. \tag{4.2}$$

Theorem 5 includes the commonly used super smooth distributions such as normal, Cauchy, and their mixtures as an error variable. Theorem 6 examines the ordinary smooth cases, which include all gamma distributions, and symmetric gamma distributions (e.g. double exponential distributions) as specific examples.

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

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

Then, for any fixed point x, *there exists a positive constant* D_3 *such that*

$$\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_{k,B,2}} E_f|\hat{T}_n(x) - m(x)|^2 > D_3 n^{-\frac{2k}{2(k+\beta)+1}}. \tag{4.3}$$

Moreover, if the weight function $w(\cdot)$ *is positive and continuous on an interval, then there exists a positive constant* D_4 *such that*

$$\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_{k,B,p}} E_f\|\hat{T}_n(\cdot) - m(\cdot)\|_{wp} > D_4 n^{-\frac{k}{2(k+\beta)+1}}, \quad \forall 1 \le p < \infty. \tag{4.4}$$

Remark 3. According to our notation, $\mathcal{F}_{k,B,p} \subset \mathcal{F}_{k,B,2}$ [see (3.3)]. Thus, lower rates with $f \in \mathcal{F}_{k,B,p}$ is stronger than lower rates with $f \in \mathcal{F}_{k,B,2}$. In fact, according to our proof, one can replace $f \in \mathcal{F}_{k,B,2}$ in (4.1)—(4.3) by $f \in \mathcal{F}_{k,B,p}$ to get a stronger result. In other words, the existence of conditional moments is not an intrinsic part of the lower rates of convergence.

The idea of establishing the lower bound is interesting and can be highlighted as follows. We use pointwise estimation (4.1) and (4.3) to illustrate the idea; the global lower bound can be treated similarly by combining the argument on pointwise estimation with the idea of adaptively local 1-dimensional subproblem of Fan (1989).

Suppose the problem is to estimate the regression function at the point 0 and that we want to obtain a lower bound for

$$\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_{k,B,2}} E_f|\hat{T}_n(0) - m(0)|^2.$$

Let $f_1(\cdot,\cdot)$ and $f_2(\cdot,\cdot)$ denote two points in $\mathcal{F}_{k,B,2}$ and put $m_1(0) = E_{f_1}(Z|X = 0)$, $m_2(0) = E_{f_2}(Z|X = 0)$. Then

$$\begin{aligned}
&\inf_{\hat{T}_n} \sup_{f\in\mathcal{F}_{k,B,2}} E_f|\hat{T}_n(0) - m(0)|^2 \\
\geq\ &\inf_{\hat{T}_n} \frac{1}{2}\left[E_{f_1}|\hat{T}_n(0) - m_1(0)|^2 + E_{f_2}|\hat{T}_n(0) - m_2(0)|^2\right] \\
\geq\ &\frac{\Delta^2}{2}\inf_{\hat{T}_n}\left[P_{f_1}(|\hat{T}_n(0) - m_1(0)| \geq \Delta) + P_{f_2}(|\hat{T}_n(0) - m_1(0)| \leq \Delta)\right], \qquad (4.5)
\end{aligned}$$

where $\Delta = |m_1(0) - m_2(0)|/2$. One can view the second factor of (4.5) as the sum of the probabilities of type I and type II errors of the best testing procedure for the problem:

$$H_0 : f(x,z) = f_1 \quad \text{vs.} \quad H_1 : f(x,z) = f_2. \qquad (4.6)$$

Now, if the testing problem (4.6) can not be tested consistently based on the data (Y_j, Z_j) ($j = 1,2,...,n$), then the second factor of (4.5) would be bounded away from 0, and hence then the difference of the functional values $|m_1(0) - m_2(0)|$ consists of a lower rate for estimating $E_f(Z|X = 0)$. See Donoho and Liu (1988) for a related idea. Thus *the problem of establishing a lower bound of nonparametric regression becomes the problem of finding a pair of densities f_1 and f_2 so that* (a) *the distance $\Delta = |m_1(0) - m_2(0)|/2$ can be computed explicitly, and* (b) *the hypotheses (4.6) can not be tested consistently.*

Finding the pair f_1 and f_2. Let f_0, g_0 denote symmetric density functions such that

$$\int_{-\infty}^{\infty} |\phi_0(t)|\, dt \leq B \qquad \text{and} \qquad \min_{a\leq x\leq b} f_0(x) \geq B^{-1}, \qquad (4.7)$$

where $\phi_0(t)$ is the characteristic function of f_0. Put $f_1(x,z) = f_0(x)g_0(z)$. Then $f_1 \in \mathcal{F}_{k,B,2}$.

Now let $h_0(x)$ be a function satisfying

$$\int_{-\infty}^{\infty} z^j h_0(z)dz = j, \quad j = 0, 1 \tag{4.8}$$

and let $\{a_n\}$ denote a sequence of positive numbers with $a_n \to 0$. Put

$$f_2(x, z) = f_0(x)g_0(z) + a_n^k H(x/a_n)h_0(z), \tag{4.9}$$

where the function $H(\cdot)$ will be specified in the proofs of Theorem 5 and 6 so that $f_2 \in \mathcal{F}_{k,B,2}$.

By (4.7) – (4.9), $m_1(0) = 0$ and $m_2(0) = a_n^k H(0)/f_0(0)$. Hence

$$|m_1(0) - m_2(0)| = a_n^k H(0)/f_0(0). \tag{4.10}$$

This yields a lower rate of estimating $E_f(Z|X = 0)$.

Recall that F_ε is the cdf of the error distribution. Let $f_j * F_\varepsilon(y, z) = \int f_j(y - x, z)\, dF_\varepsilon(x)$ $(j = 1, 2)$ denote the densities of (Y, Z) under H_0 and H_1, respectively. The hypotheses (4.6) can not be tested consistently if $f_1 * F_\varepsilon$ is too close to $f_2 * F_\varepsilon$. In terms of the χ^2-distance, this is given by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(f_1 * F_\varepsilon - f_2 * F_\varepsilon)^2}{f_1 * F_\varepsilon} dydz = O\left(\frac{1}{n}\right). \tag{4.11}$$

One advantage of using χ^2-distance is that it can be readily approximated by the L^2 distance so that Parseval's identity can be used. See Fan (1988) for further discussions on using various distances in obtaining lower bounds in deconvolution problems.

Note that $[(f_1 - f_2) * F_\varepsilon](y, z) = a_n^k h_0(z)[H(\cdot/a_n) * F_\varepsilon(y)]$ and $f_1 * F_\varepsilon(y, z) = g_0(z)[f_0 * F_\varepsilon(y)]$. By (4.11),

$$a_n^{2k} \int_{-\infty}^{\infty} \frac{h_0^2}{g_0} dz \left(\int_{-\infty}^{\infty} \frac{[H(\cdot/a_n) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy \right) = O\left(\frac{1}{n}\right). \tag{4.12}$$

Combining (4.10) and (4.12), the lower rate of estimating $m(0)$ is a_n^k with a_n satisfying

$$a_n^{2k} \int_{-\infty}^{\infty} \frac{[H(\cdot/a_n) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy = O\left(\frac{1}{n}\right). \tag{4.13}$$

If we use the same argument for estimating a density f_X at the point 0 from the convolution model $Y = X + \varepsilon$, we will end up exactly the same problem: finding a_n from (4.13).

See Section 3 of Fan (1988). Now for the deconvolution problem, the solution of a_n is found by Fan (1988). By the result of Fan (1988), we get the desired lower rates for estimating the regression function at a point.

Let's illustrate how a_n is related to the tail behavior of the characteristic function of the error variable. By suitable choice of f_0, H, it can be shown that the χ^2-distance is equivalent to the L_2 distance:

$$\int_{-\infty}^{\infty} \frac{[H(\cdot/a_n) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy = O\left(\int_{-\infty}^{\infty} [H(\cdot/a_n) * F_\varepsilon]^2\, dy\right). \tag{4.14}$$

By Parseval's identity,

$$\begin{aligned}\int_{-\infty}^{\infty} [H(\cdot/a_n) * F_\varepsilon]^2\, dy &= \frac{1}{2\pi}\int_{-\infty}^{\infty} |a_n \phi_H(a_n t)\phi_\varepsilon(t)|^2\, dt \\ &= \frac{a_n}{2\pi}\int_{-\infty}^{\infty} |\phi_H(t)\phi_\varepsilon(t/a_n)|^2\, dt. \end{aligned}\tag{4.15}$$

If $\phi_H(t)$ vanishes when $|t| \le 1$, then (4.15) depends only on the tail of $\phi_\varepsilon(t)$. In other words, we would choose a_n so that

$$a_n^{2k+1}\int_{|t|\ge 1} |\phi_H(t)\phi_\varepsilon(t/a_n)|^2\, dt = O\left(\frac{1}{n}\right), \tag{4.16}$$

and it is easy to do so by our assumption of the tail of characteristic functions.

One final remark: our method of perturbation is quite different from those in the literature of nonparametric regression (see e.g. Stone (1980, 1982) and 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 can *not* handle our more sophisticated error-in-variable problems. We should also mention that the connection has other applications as well. For example, it can be used to determine the minimax risk of nonparametric regression, by connecting the risk with the minimax risk of estimating a nonparametric density.

5 Concluding Remarks

Nonparametric regression has been very popular because of its flexibility in fitting a large variety of data sets. However, this method breaks down in situations when covariates are measured with errors. To remedy this, the current approach proposes a new method in nonparametric estimation of regression function with errors-in-variables. This is accomplished by modifying the usual kernel method so that deconvolution is involved to provide an estimate of the marginal density for the unobserved covariates. We then examine the effects of errors-in-variables on the modified regression estimators. It is shown that the current approach possesses various optimal properties depending the type of error distributions, and a lot of insights have been gained in this investigation. Some of these are highlighted as follows:

- The difficulty of nonparametric regression with errors-in-variables depends strongly on the smoothness of error distribution: the smoother, the harder. This provides a new understanding of intrinsic features of the problems, which is expected to have other applications such as "ill-posed" problems.

- As opposed to the approach to regression analysis with errors-in-variables based on normal error distributions, our study shows that this popular method suffers the draw back that the kernel estimators have extremely slow rates of convergence. We also show that this is the intrinsic part of the problem and is not an artifact of the kernel method.

- 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.

- 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.

6 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 $Y = X+\varepsilon$,

$$g(y,z) = \int f(y-x,z)\,dF_\varepsilon(x), \tag{6.1}$$

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

6.1 Proof of Theorem 1.

The proof of this theorem depends on the following lemmas.

Lemma 6.1. *Under the conditions of Theorem 1,*

$$E\sup_x |\hat{f}_n(x) - f_X(x)|^p = o(1), \tag{6.2}$$

where $\hat{f}_n$ is defined by (2.5).

Proof. Let $\phi_X(t)$ be the characteristic function of X. By (2.3),

$$\sup_x |\hat{f}_n(x) - f_X(x)| \le \frac{1}{2\pi}\int_{-\infty}^{\infty} |\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_Y(t)} - 1||\phi_X(t)|\,dt. \tag{6.3}$$

Then

$$\begin{aligned} E[\sup_x |\hat{f}_n(x) - f_X(x)|]^p &\le \left(\frac{1}{2\pi}\right)^p E\left(\int_{-\infty}^{\infty} |\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_Y(t)} - 1||\phi_X(t)|\,dt\right)^p \\ &\le \left(\frac{1}{\pi}\right)^p E\left(\int_{-\infty}^{\infty} \frac{|\phi_X(t)||\phi_K(th_n)|}{|\phi_Y(t)|}|\hat{\phi}_n(t) - \phi_Y(t)|\,dt\right)^p \\ &\quad + \left(\frac{1}{\pi}\right)^p \left(\int_{-\infty}^{\infty} |\phi_K(th_n) - 1||\phi_X(t)|\,dt\right)^p. \end{aligned} \tag{6.4}$$

It follows from Hölder's inequality and the fact that there exists a constant c_p such that

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

that

$$\begin{aligned}
& E\left(\int_{-\infty}^{\infty} \frac{|\phi_X(t)||\phi_K(th_n)|}{|\phi_Y(t)|}|\hat{\phi}_n(t) - \phi_Y(t)|\,dt\right)^p \\
\leq\ & \left(\int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}\,dt\right)^{p-1} \int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|} E[|\hat{\phi}_n(t) - \phi_Y(t)|^p]\,dt \\
\leq\ & \left(\int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\varepsilon(t)|}\,dt\right)^p \cdot c_p n^{-p/2} \\
=\ & c_p \left(\frac{1}{\sqrt{n}h_n} \int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt\right)^p. \qquad (6.5)
\end{aligned}$$

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 $\phi_K(t)$,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt &= 2\int_0^{M_0} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt \\
&\leq 2\int_0^{Mh_n} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt + \frac{4}{d_0}\int_{Mh_n}^{M_0} |\phi_K(t)| \left|\frac{t}{h_n}\right|^{-\beta_0} \exp(|t/h_n|^\beta/\gamma)\,dt \\
&= O(h_n) + O\left(h_n^{-1} \exp(|M_0/h_n|^\beta/\gamma)\right). \qquad (6.6)
\end{aligned}$$

Combining (6.4)—(6.6), the conclusion follows from the choice of h_n.

Lemma 6.2. Let $A_n(x) = (nh_n)^{-1}\sum K_n((x - Y_j)/h_n)[Z_j - m(x)]$. If $\phi_K(\cdot)$ vanishes outside the interval $[-M_0, M_0]$, then

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

Proof. According to (6.1) and (2.6),

$$\begin{aligned}
h_n EA_n(x) &= EK_n((x - Y_1)/h_n)[Z_1 - m(x)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_n((x-y)/h_n)[z - m(x)]g(y,z)\,dydz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_n(\frac{x-y}{h_n})[z - m(x)]f(y-u,z)\,dF_\varepsilon(u)dydz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi}\exp(-it(x-y)/h_n)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} \\
&\qquad \times [z - m(x)]f(y-u,z)\,dtdF_\varepsilon(u)dydz. \qquad (6.7)
\end{aligned}$$

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

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

Thus

$$\begin{aligned}
& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi h_n} \exp(-it(x-y)/h_n) \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} f(y-u,z)\, dF_\varepsilon(u) dy dt \\
= \quad & \int_{-\infty}^{\infty} \frac{1}{2\pi h_n} \exp(-itx/h_n) \frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)} \cdot \left\{ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(ity/h_n) f(y-u,z)\, dF_\varepsilon(u) dy \right\} dt \\
= \quad & \int_{-\infty}^{\infty} \frac{1}{2\pi h_n} \exp(-itx/h_n) \phi_K(t) \left\{ \int_{-\infty}^{\infty} \exp(ity/h_n) f(y,z)\, dy \right\} dt \\
= \quad & \frac{1}{h_n} \int_{-\infty}^{\infty} K(\frac{x-y}{h_n}) f(y,z)\, dy, \qquad (6.8)
\end{aligned}$$

here the last equality follows from the inversion of Fourier transform: the inversion of the products of two Fourier transforms equals to the convolution. By (6.7) and (6.8),

$$\begin{aligned}
EA_n(x) \quad &= \quad \frac{1}{h_n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [z - m(x)] K(\frac{x-y}{h_n}) f(y,z)\, dy dz \\
&= \quad \frac{1}{h_n} \int_{-\infty}^{\infty} [m(u) - m(x)] K(\frac{x-u}{h_n}) f_X(u)\, du.
\end{aligned}$$

Proof of Theorem 1. According to Lemma 6.1, $\hat{f}_n(x) = f_X(x) + [\hat{f}_n(x) - f_X(x)]$ is bounded from below in probability. By linearization,

$$\begin{aligned}
\hat{m}_n(x) - m(x) \quad &= \quad [\hat{m}_n(x) - m(x)] \frac{\hat{f}_n(x)}{f_X(x)} (1 + \xi_n(x)) \\
&= \quad \frac{n^{-1} \sum h_n^{-1} K_n((x - Y_j)/h_n)[Z_j - m(x)]}{f_X(x)} (1 + \xi_n(x)), \qquad (6.9)
\end{aligned}$$

where $\xi_n(x) = [f_X(x) - \hat{f}_n(x)]/\hat{f}_n(x)$. By Lemma 6.1, $E|\xi_n(x)|^2 = o(1)$. Thus, the leading term is $A_n(x)/f_X(x)$, which by Lemma 6.2 has "bias"

$$\begin{aligned}
EA_n(x) \quad &= \quad \frac{1}{h_n} \int_{-\infty}^{\infty} [m(u) - m(x)] K(\frac{x-u}{h_n}) f_X(u)\, du \\
&= \quad f_X(x) b_k(x) h_n^k (1 + o(1)). \qquad (6.10)
\end{aligned}$$

Let $\sigma^2(x) = E((Z_1 - m(x))^2 | X = x)$. Since X and ε are independent,

$$h_n^2 \mathrm{Var}(A_n(x)) \quad = \quad \frac{1}{n} \mathrm{Var}\left(K_n(\frac{x-Y}{h_n})[Z - m(x)] \right)$$

$$\begin{aligned} &\leq \frac{1}{n} E|K_n\left(\frac{x-Y}{h_n}\right)|^2[Z-m(x)]^2 \\ &= \frac{1}{n} E K_n^2\left(\frac{x-X-\varepsilon}{h_n}\right)\sigma^2(X). \end{aligned} \tag{6.11}$$

By (2.6) and (6.6),

$$\sup_x |K_n(x)| = O\left(\exp(|M_0/h_n|^\beta/\gamma)/h_n\right).$$

It follows from (6.11) and $h_n = c(\log n)^{-1/\beta}$ that

$$\mathrm{Var}(A_n(x)) = O\left(\frac{1}{nh_n^3}\exp(2|M_0/h_n|^\beta/\gamma)\right) E\sigma^2(X) = o(h_n^{2k}). \tag{6.12}$$

Eq. (3.4) now follows from the usual bias and variance decomposition. Since (6.10) and (6.12) hold uniformly in $x \in (a,b)$, the second conclusion is also valid.

6.2 Proof of Theorem 2.

Note that (6.2) also holds uniformly in the class of $f \in \mathcal{F}_{k,B,2}$. That is,

$$\sup_{f\in\mathcal{F}_{k,B,2}} E\sup_x |\hat{f}_n(x) - f_X(x)|^p = o(1). \tag{6.13}$$

Thus, by the linearization argument (6.9), we need only to argue that

$$\begin{aligned} \sup_{f\in\mathcal{F}_{k,B,2}} E\sup_x |A_n(x)|^2 &\leq 2\sup_{f\in\mathcal{F}_{k,B,2}}\sup_x |EA_n(x)|^2 + 2\sup_{f\in\mathcal{F}_{k,B,2}} E\sup_x |A_n(x) - EA_n(x)|^2 \\ &= O((\log n)^{-2k/\beta}), \end{aligned} \tag{6.14}$$

where

$$A_n(x) = n^{-1}\sum_j h_n^{-1} K_n((x-Y_j)/h_n)[Z_j - m(x)] \tag{6.15}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{1}{n}\sum_{j=1}^{n}\exp(itY_j)(Z_j - m(x))\frac{\phi_K(th_n)}{\phi_\varepsilon(t)}dt. \tag{6.16}$$

Note that (6.14) has the usual bias-variance decomposition. By (6.10),

$$\sup_{f\in\mathcal{F}_{k,B,2}}\sup_x |EA_n(x)|^2 = O(h_n^{2k}). \tag{6.17}$$

Thus, we need only to argue that the "variance" term in (6.14) has the right order. Put

$$U_j(t) = \exp(itY_j)Z_j - E\exp(itY_j)Z_j \quad \text{and} \quad V_j(t) = \exp(itY_j) - E\exp(itY_j).$$

Then, according to (6.16) and the fact that $|m(x)| \leq B$,

$$\begin{aligned}
&\left(2\pi \sup_x |A_n(x) - EA_n(x)|\right)^2 \\
\leq\ & \left[\int_{-\infty}^{\infty} \left(|\frac{1}{n}\sum_{j=1}^n U_j(t)| + \frac{B}{n}|\sum_{j=1}^n V_j(t)|\right) \left|\frac{\phi_K(th_n)}{\phi_\epsilon(t)}\right| dt\right]^2.
\end{aligned}$$

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

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\epsilon(t)|} dt \int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\epsilon(t)|}\left[|\frac{1}{n}\sum_{j=1}^n U_j(t)| + |\frac{B}{n}\sum_{j=1}^n V_j(t)|\right]^2 dt \\
\leq\ & 2\int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\epsilon(t)|} dt \int_{-\infty}^{\infty} \frac{|\phi_K(th_n)|}{|\phi_\epsilon(t)|}\left[|\frac{1}{n}\sum_{j=1}^n U_j(t)|^2 + |\frac{B}{n}\sum_{j=1}^n V_j(t)|^2\right] dt. \qquad (6.18)
\end{aligned}$$

Note that

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

By (6.18),

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

By (6.6) 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.

6.3 Proof of Theorem 3.

The proof uses the previous linearization argument, which in turn depends on the following result for ordinary smooth models.

Lemma 6.3. *Under the conditions of Theorem 3, if $h_n \to 0$ and $nh^{2\beta+2} \to \infty$, then*

$$E\sup_x |\hat{f}_n(x) - f_X(x)|^p = o(1), \qquad (6.19)$$

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

Proof. By (6.4) and (6.5), it is sufficient to verify that

$$\int_{-\infty}^{\infty} \frac{\phi_K(t)}{\phi_\epsilon(t/h_n)} dt = o(\sqrt{n}h_n). \qquad (6.20)$$

In fact, according to (3.8), there are positive constants M and c_1 such that

$$|\phi_\varepsilon(t)| \geq c_1 |t|^{-\beta} \text{ for } |t| > M. \tag{6.21}$$

Therefore

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt &= 2\int_0^{\infty} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt \\
&\leq 2\int_0^{Mh_n} \frac{|\phi_K(t)|}{|\phi_\varepsilon(t/h_n)|}\,dt + 2\int_{Mh_n}^{\infty} \frac{|\phi_K(t)|}{c_1} \left|\frac{t}{h_n}\right|^{\beta} dt \\
&\leq 2Mh_n \frac{\max|\phi_K(t)|}{\min|\phi_\varepsilon(t)|} + \frac{2h_n^{-\beta}}{c_1} \int_0^{\infty} |\phi_K(t)| t^{\beta}\,dt \\
&= O(h_n^{-\beta}).
\end{aligned} \tag{6.22}$$

Proof of Theorem 3. It follows from Lemma 6.3 that the linearization argument given in the proof of Theorem 1 also holds in the ordinary smooth models. Thus, to prove the theorem, it suffices to compute the bias and the variance of

$$A_n(x) = n^{-1} \sum_j h_n^{-1} K_n((x - Y_j)/h_n)[Z_j - m(x)]. \tag{6.23}$$

Since the bias of $A_n(x)$ does not depend on the error distribution (see Lemma 6.2), it follows from (6.10) that

$$EA_n(x) = f_X(x) b_k(x) h_n^k (1 + o(1)). \tag{6.24}$$

The variance is given by

$$\begin{aligned}
\mathrm{Var}(A_n(x)) &= \frac{1}{n}\mathrm{Var}\left(h_n^{-1} K_n(\frac{x-Y}{h_n})[Z - m(x)]\right) \\
&= \frac{1}{n} E|h_n^{-1} K_n\left(\frac{x-Y}{h_n}\right)|^2 [Z - m(x)]^2 + o(\frac{1}{n}) \\
&= \frac{1}{n} E|h_n^{-1} K_n\left(\frac{x-X-\varepsilon}{h_n}\right)|^2 \sigma^2(X) + o(\frac{1}{n}) \\
&= \frac{1}{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |h_n^{-1} K_n\left(\frac{x-u-v}{h_n}\right)|^2 \sigma^2(u) f_X(u)\,dF_\varepsilon(v)\,du + o(\frac{1}{n}) \\
&= \frac{1}{nh_n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_n^2(u) \sigma^2(x - v - uh_n) f_X(x - v - uh_n)\,dF_\varepsilon(v)\,du \\
&\quad + o(\frac{1}{n}).
\end{aligned} \tag{6.25}$$

To this end, note that by (3.8) and the dominated convergence theorem

$$h_n^{\beta} K_n(y) \to \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-ity) \frac{t^{\beta}}{c} \phi_K(t)\, dt \stackrel{\text{def}}{=} J(y). \tag{6.26}$$

By Lemma 6.4 (to be given at the end of this section),

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

for some positive constant C. According to (6.25) and Lemma 2.1 of Fan (1990),

$$\mathrm{Var}(A_n(x)) = \frac{1}{nh_n^{1+2\beta}} \int_{-\infty}^{\infty} J^2(u)\, du \int_{-\infty}^{\infty} \sigma^2(x - v) f_X(x - v)\, dF_{\varepsilon}(v)[1 + o(1)]. \tag{6.27}$$

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

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

The conclusion follows from bias and the variance decomposition.

Lemma 6.4. *Under the conditions of Theorem 3,*

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

for some constant C.

Proof. By the definition of $K_n(x)$ and (6.22),

$$|h_n^{\beta} K_n(y)| \leq \frac{h_n^{\beta}}{2\pi} \int_{-\infty}^{\infty} \frac{|\phi_K(t)|}{|\phi_{\varepsilon}(t/h_n)|} dt = O(1). \tag{6.28}$$

Now, by integration by parts, and the similar virtue as (6.22),

$$|h_n^{\beta} K_n(y)| \leq \frac{h_n^{\beta}}{2\pi |y|} \int_{-\infty}^{\infty} \left| \left(\frac{\phi_K(t)}{\phi_{\varepsilon}(t/h_n)} \right)' \right| dt \leq D|y|^{-1}, \tag{6.29}$$

where D is a positive constant. The desired conclusion follows from (6.28) and (6.29).

6.4 Proof of Theorem 4

The local rate (3.11) follows the similar argument given in Theorem 3. We focus on proving the global results (3.12):

$$\sup_{f\in\mathcal{F}_{k,B,p}} E\|\hat{m}_n(x)-m(x)\|_{wp}^p = O\left(n^{-\frac{pk}{2(k+\beta)+1}}\right). \tag{6.30}$$

If we show that [see (6.23) for the definition of $A_n(x)$]

$$\sup_{f\in\mathcal{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)dx = O\left(n^{-\frac{pk}{2(k+\beta)+1}}\right), \tag{6.31}$$

then by linearization (6.9), we have

$$\begin{aligned}
\sup_{f\in\mathcal{F}_{k,B,p}} E\|\hat{m}_n(x)-m(x)\|_{wp}^p &= \sup_{f\in\mathcal{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)\left[(1+\xi_n(x))/f_X(x)\right]^p dx \\
&\le B^p \sup_{f\in\mathcal{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)dx(1+o(1)) \\
&= O\left(n^{-\frac{pk}{2(k+\beta)+1}}\right).
\end{aligned}$$

The middle inequality follows from the fact that Lemma 6.3 also holds uniformly in $f\in\mathcal{F}_{k,B,p}$.

Now, let's turn to prove (6.31). By Lemma 6.2 and (6.24)

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

Thus,

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_{k,B,p}} E\int_a^b |A_n(x)|^p w(x)dx \\
\le\ & 2^p \sup_{f\in\mathcal{F}_{k,B,p}} E\int_a^b |A_n(x)-EA_n(x)|^p w(x)dx + O\left(n^{-\frac{pk}{2(k+\beta)+1}}\right).
\end{aligned} \tag{6.32}$$

Hence, we need only to justify that the first term of (6.32) is of the right order $O\left(n^{-\frac{pk}{2(k+\beta)+1}}\right)$.

Let r be the smallest integer exceeding $p/2$ and put

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

Then

$$E|A_n(x) - EA_n(x)|^p = E|\frac{1}{n}\sum_{j=1}^{n} T_j(x)|^p \le \left(E|\frac{1}{n}\sum_{j=1}^{n} T_j(x)|^{2r}\right)^{\frac{p}{2r}}. \tag{6.33}$$

Moreover,

$$\sup_{f\in\mathcal{F}_{k,B,p}} \sup_{x\in[a,b]} E\left(\frac{1}{n}\sum_{j=1}^{n} T_j(x)\right)^{2r} = O\left((\frac{1}{n}h_n^{1-2(1+\beta)})^r\right). \tag{6.34}$$

[Proof of (6.34) will be given shortly.] The conclusion of the theorem follows from (6.32)—(6.34).

We now prove (6.34) by a pair of lemmas, which hold uniformly in $f \in \mathcal{F}_{k,B,p}$.

Lemma 6.5. *Under the conditions of Theorem 4,*

$$\sup_x E|T_1(x)|^l = O\left(h_n^{1-l(\beta+1)}\right) \text{ for } j = 2, \ldots, r. \tag{6.35}$$

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

$$\begin{aligned}
h_n^l E|T_1(x)|^l &\le 2^{l+1}E\left|K_n((x-Y_1)/h_n)[Z_1 - m(x)]\right|^l \\
&\le 2^{2l+1}\left[E|K_n((x-Y)/h_n)Z|^l + B^l E|K_n((x-Y)/h_n)|^l\right] \\
&= 2^{2l+1}\left[E|K_n((x-X-\varepsilon)/h_n)|^l\nu_l(X) + B^l E|K_n((x-Y)/h_n)|^l\right] \\
&\le 2^{2l+1}(B+B^l)E|K_n((x-X-\varepsilon)/h_n)|^l. \qquad (6.36)
\end{aligned}$$

Recall that $f_Y(y)$ is the density of $Y = X + \varepsilon$. Then by Lemma 6.4,

$$\begin{aligned}
E|K_n((x-X-\varepsilon)/h_n)|^l &\le h_n\int_{-\infty}^{\infty}|K_n(y)|^l f_Y(x-yh_n)\,dy \\
&\le C^l h_n^{1-l\beta}\int_{-\infty}^{\infty}\frac{1}{(1+|y|)^l}f_Y(x-yh_n)\,dy \\
&= O\left(h_n^{1-l\beta}\right). \qquad (6.37)
\end{aligned}$$

The desired result follows from (6.36) and (6.37).

Lemma 6.6. *Under the conditions of Lemma 6.5,*

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

Proof. Write $T_j = T_j(x)$. By the multinomial 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 $\sum'$ sums over k-tuples of positive integers $(r_1, \ldots, r_k)$ satisfying $r_1 + \cdots + r_k = 2r$ and $\sum''$ extends over k-tuples of distinct integers $(j_1, \ldots, j_k)$ in the range $1 \leq j \leq n$.

By independence and that T_j has mean zero,

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

where $\sum'''$ sums over k-tuples of positive integers $(r_1, \ldots, r_k)$ satisfying $r_1 + \cdots + r_k = 2r$ and $r_j \geq 2$ $(j = 1, \ldots, k)$. Thus $k \leq r$. By Lemma 6.5,

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

since $nh_n \to \infty$. The desired result follows. This completes the proof of Lemma 6.6.

6.5 Proof of Theorem 5 & 6

We first justify the local lower rates of Theorems 5 and 6, i.e. (4.1) and (4.3). The basic idea is outlined in Section 4. For simplicity, we prove only for the case $x = 0$ in (4.1) and (4.3).

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

$$f_0(x) = C_r(1 + x^2)^{-r}, (r > 0.5) \qquad g_0(z) = \left(\sqrt{2\pi}b\right)^{-1} \exp(-\frac{z^2}{2b^2}) \tag{6.38}$$

and

$$h_0(z) = \frac{1}{\sqrt{2\pi}} \left(\exp(-(z-1)^2) - \exp(-z^2)\right), \tag{6.39}$$

where $C_r = \int_{-\infty}^{\infty}(1+x^2)^{-r}dx$, and b, r will be chosen later. Note that h_0 satisfies (4.8). To construct the function H, note that the heuristic argument suggests that the Fourier transform $\phi_H(t)$ of the function H should vanish for $|t| \leq 1$, and in order to make the integrability of (4.15), it is imposed that $\phi_H(\cdot)$ vanishes when $|t| \geq 2$. (Note that the interval [1, 2] is chosen for convenience.) Now let's show how to construct such a function.

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 following properties:

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

By the proper choice of r, b, the pair of densities f_1, f_2 defined by (4.9) will be members of $\mathcal{F}_{k,B,2}$. By the argument in Section 4, a_n^k would be the lower rates if a_n satisfies (4.13). According to Fan (1988) and the conditions of Theorem 5, there is a positive constant c_1 such that the solution of a_n to (4.13) is given by

$$a_n = (\log n + c_1 \log(\log n))^{-1/\beta}\gamma^{-1/\beta}. \tag{6.40}$$

Similarly, for Theorem 6, the solution is given by

$$a_n = c_2 n^{-1/(2k+2\beta+1)}, \tag{6.41}$$

where c_2 is a positive constant. Thus, the conclusion of (4.1) and (4.3) follows.

Now, let's turn to the global rates (4.2) and (4.4). We use the idea of adaptively local 1-dimension subproblem of Fan (1989). 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 infinity and $x_j = j/m_n$ $(j = 1, 2, ..., m_n)$ be a grid point of $[0,1]$. Let $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_{m_n}) \in \{0,1\}^{m_n}$ be a vector of 0, and 1. Construct a sequence of functions

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

Define a family of densities:

$$f_{\boldsymbol{\theta}}(x,z) = f_0(x)g_0(z) + \delta m_{\boldsymbol{\theta}}(x)h_0(z), \tag{6.42}$$

where f_0, g_0, and h_0 is defined by (6.38) and (6.39).

For suitable choice of $b > 0, r > 0.5$ and $\delta > 0$, we now show that $f_{\boldsymbol{\theta}}(x,z) \in \mathcal{F}_{k,B,p}$, which is a subset of $\mathcal{F}_{k,B,2}$. It is easy to see that $|h_0(z)| \le c_3 g_0(z)$ for all z, and by Lemma 6.7, $|m_{\boldsymbol{\theta}}(x)| \le c_4(1 + x^2)^{-m_0/2}$. Thus, for sufficiently small $\delta > 0$ and $r - 0.5 > 0$, $f_{\boldsymbol{\theta}}$ is a density function satisfying

$$f_{\boldsymbol{\theta}}(x,z) \ge 0.5 f_0(x)g_0(z), \qquad \forall \boldsymbol{\theta} \in \{0,1\}^{m_n}. \tag{6.43}$$

Now, the conditional mean is given by

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

By Lemma 6.7 again, the kth derivative of the conditional expectation is bounded by the constant B for small $\delta > 0$. Similarly, the conditions of conditional moments and the marginal density are satisfied for suitable choice of $r > 0.5$ and $\delta > 0$. Hence, $f_{\boldsymbol{\theta}}(x,z) \in \mathcal{F}_{k,B,p}$ for all $\boldsymbol{\theta} \in \{0,1\}^{m_n}$.

Denote

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

and

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

Then there is a positive constant c_5 so that the difference of functional values for the pair of densities satisfies

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

Put

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

By Theorem 1 of Fan (1989), if

$$\max_{1 \leq j \leq m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2 \left(f_{\boldsymbol{\theta}_{jo}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon \right) \leq c_6/n, \tag{6.44}$$

then

$$\begin{aligned} & \inf_{\hat{T}_n(x)} \sup_{f \in \mathcal{F}_{k,B,p}} E_f \int_0^1 |\hat{T}_n(x) - m(x)|^p w(x) dx \\ \geq\ & \frac{1 - \sqrt{1 - \exp(-c_6)}}{2^{p+1}} \int_0^1 w(x) dx \int_0^1 |H(x)|^p dx (c_5 m_n^{-k})^p. \end{aligned} \tag{6.45}$$

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

Now, let's determine m_n from (6.44). By (6.43), we have

$$\begin{aligned} & \max_{1 \leq j \leq m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2 \left(f_{\boldsymbol{\theta}_{jo}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon \right) \\ \leq\ & 2 m_n^{-2k} \int_{-\infty}^{\infty} h_0^2(z)/g_0(z) dz \int_{-\infty}^{\infty} \frac{[H(m_n(\cdot - x_j)) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dy \end{aligned} \tag{6.46}$$

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 the denominator of (6.46) with a change of variable, we have

$$\begin{aligned} & \max_{1 \leq j \leq m_n} \max_{\theta \in \{0,1\}^{m_n}} \chi^2 \left(f_{\boldsymbol{\theta}_{jo}} * F_\varepsilon, f_{\boldsymbol{\theta}_{j1}} * F_\varepsilon \right) \\ \leq\ & 2 m_n^{-2k} c_7^{-1} \int_{-\infty}^{\infty} h_0^2(z)/g_0(z) dz \quad \int_{-\infty}^{\infty} \frac{[H(m_n(\cdot)) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dx. \end{aligned} \tag{6.47}$$

In other words, we need to determine m_n from the equation

$$m_n^{-2k} \int_{-\infty}^{\infty} \frac{[H(m_n(\cdot)) * F_\varepsilon]^2}{f_0 * F_\varepsilon} dx = O\left(\frac{1}{n}\right). \tag{6.48}$$

The problem (6.48) is exactly the same as problem (4.13), by thinking of $a_n = m_n^{-1}$. The conclusion follows again from Fan (1988) [see also (6.40) and (6.41)].

Lemma 6.7. *Suppose that the function $G(x)$ satisfies*

$$|G(x)| \le C(1+x^2)^{-m} \quad (m > 0.5). \tag{6.49}$$

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

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

Proof. If $|x| \ge 2$, then there is a positive constant C_1 such that

$$\begin{aligned}
\sum_{j=1}^{m_n} |G(m_n x - j)| &\le C \sum_{j=1}^{m_n} \frac{1}{(m_n x^2 - m_n)^m} \\
&= C m_n^{1-2m} (x^2 - 1)^{-m} \\
&\le C_1 (1+x^2)^{-m}.
\end{aligned}$$

When $|x| < 2$,

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

as have to be shown.

References

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

[2] Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the errors in variables model. *Ann. Statist.*, **15**, 513–540.

[3] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density. *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

[4] 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*, **70**, 19–25.

[5] Donoho, D. and Liu, R. (1988). Geometrizing rates of Convergence I, *Tech. Report.* 137, Dept. of Statistics, University of California, Berkeley.

[6] Fan, J. (1988). On the optimal rates of convergence for nonparametric deconvolution problem. *Tech. Report 157*, Dept. of Statistics, University of California, Berkeley.

[7] Fan, J. (1989). Adaptively local 1-dimensional subproblems. *Institute of Statistics Mimeo Series #2010*, Univ. of North Carolina, Chapel Hill.

[8] Fan, J. (1990). Asymptotic normality for deconvolving kernel density estimators. To appear in *Sankhyā.*

[9] Fuller, W. A. (1986). *Measurement error models.* Wiley, New York.

[10] Liu, M. C. and Taylor, R. L. (1990). A consistent nonparametric density estimator for the deconvolution problem. To appear in *Canadian J. Statist.*

[11] 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.

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

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

[14] Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.*, **13**, 1335–1351.

[15] Stefanski, L. A. and Carroll, R. J. (1987a). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703–716.

[16] Stefanski, L. A. and Carroll, R. J. (1987b). Deconvoluting kernel density estimators. *Institute of Statistics Mimeo Series #1623*, Univ. of North Carolina, Chapel Hill.

[17] Stone, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, **8**, 1348-1360.

[18] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040-1053.

[19] Whittemore, A.S. and Keller, J.B. (1986). Approximations for errors in variables regression. *J. Amer. Statist. Assoc.*, **83**, 1057-1066.

[20] Whittemore, A. S. (1989). Errors-in-variables regression using Stein estimates. To appear in *American Statistician.*

[21] Zhang, C. H. (1989). Fourier methods for estimating mixing densities and distributions. *Manuscript* .