NONPARAMETRIC FUNCTION ESTIMATION INVOLVING 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

Yonghua Wang
Department of Statistics & Biostatistics
University of North Carolina
Chapel Hill, N.C. 27514

September 16, 1990

Abstract

We examine the effect of errors in covariates in nonparametric function estimation. These functions include densities, regressions and conditional quantiles. To estimate these functions, we use the idea of deconvoluting kernels in conjunction with the ordinary kernel methods. We also discuss a new class of function estimators based on local polynomials.

---

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

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

Key words and phrases. Nonparametric regression; Deconvolution; Kernel estimator; Errors in variables; Conditional quantile.

1 Introduction

In nonparametric regression analysis, one is interested in analyzing the relationship between the response Y and the covariate X through the regression function $E(Y|X)$. Suppose now X contains another random variable X° so that $X = X^\circ + \varepsilon$—a situation in which the variable of interest (X°) is *measured with error*. An important issue arises as how to recover the association between Y and X°. Note that X° is not directly observable and this complicates the problem of estimating the regression function $E(Y|X^\circ)$ since one will have to "recover" X° from X first. Alternately, it may be worth exploring the association between Y and X° by considering the conditional median function $\text{med}(Y|X^\circ)$, especially in situations involving asymmetric conditional distributions. Note that it is necessary to make this problem identifiable by assuming the "error" ε has a *known* distribution and that it is *independent* of X° and Y.

Given a training set $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the distribution of (X, Y) with $X = X^\circ + \varepsilon$, the problem of estimating the regression function $m(X^\circ) = E(Y|X^\circ)$ or the conditional median function $m(X^\circ) = \text{med}(Y|X^\circ)$ is called regression analysis with errors-in-variables. It is said to be *parametric* if the regression function is assumed to be a specific function with unknown parameters. See, for example, Armstrong (1985), Stefanski (1985), Stefanski and Carroll(1985), Prentice (1986), Whittemore and Keller (1986) and Fuller (1987). In this approach, there is no formal ways of verifying the appropriateness of the regression model. Moreover, the parameters are estimated by maximizing the likelihood equation which is usually very complicated.

To overcome the above difficulties in parametric analysis, the present paper adopts the nonparametric approach by estimating the regression function directly. In the absence of measurement errors, there is now a hugh literature on regression function estimation for studying the structures between Y and X. In situation with measurement errors, it is necessary to modify these methods since regression is used to study the effect of X° on Y, not X on Y. To achieve this, we must know how to recover X° from X. This operation is called *deconvolution* and it was first used in density estimation involving measurement errors.

This paper is outlined as follows. Section 2 begins with the problem of density estimation based on deconvoluted kernel method, since it is really the building block for estimating other functions. It then considers the problem of estimating nonparametric regression functions such as conditional mean and conditional median. Section 3 reviews some optimal properties possessed by these estimators. Section 4 discusses generalizations of these kernel methods to design-adaptive method based on local polynomials. Numerical examples are given in Section 5. Section 6 contains some concluding remarks.

2 Methods

2.1 Deconvoluted kernel density estimators

Set $X = X^\circ + \varepsilon$ and let $X_1, \ldots, X_n$ denote a training sample from the distribution of X. Assume that the error ε has a non-vanishing characteristic function $\phi_\varepsilon(t) \neq 0$. Denote the characteristic functions of X and X° by $\phi_X(\cdot)$ and $\phi_{X^\circ}(\cdot)$, respectively. By Fourier inversion, the density function of X° is given by

$$f_{X^\circ}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_{X^\circ}(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\frac{\phi_X(t)}{\phi_\varepsilon(t)}\, dt.$$

Thus the problem of estimating the density function of X° can be reduced to the estimation of the function

$$\phi_X(t) = \int_{-\infty}^{\infty} \exp(itx) f_X(x)\, dx,$$

which, in turn is a density estimation problem based on X. To this end, let $K(\cdot)$ denote a kernel function and h_n be a bandwidth. Suppose the density function $f_X(\cdot)$ of X (no errors) is now estimated by the usual kernel method:

$$\hat{f}(x) = \frac{1}{nh_n} \sum_1^n K\left(\frac{x - X_j}{h_n}\right).$$

Then we arrive at an estimator of $\phi_X(\cdot)$:

$$\begin{aligned}
\hat{\phi}_X(t) &= \int_{-\infty}^{\infty} \exp(itx)\hat{f}(x)\, dx \\
&= \int_{-\infty}^{\infty} \exp(itx)\frac{1}{nh_n} \sum_1^n K\left(\frac{x - X_j}{h_n}\right)\, dx \\
&= \frac{1}{n} \sum_1^n \exp(itX_j) \int_{-\infty}^{\infty} \exp(ituh_n)K(u)\, du \\
&= \hat{\phi}_n(t)\phi_K(th_n),
\end{aligned}$$

where $\phi_K(\cdot)$ is the Fourier transform of $K(\cdot)$ and $\hat{\phi}_n(t) = n^{-1}\sum_1^n e^{itX_j}$. This leads to a natural estimator of the density function of X°:

$$\begin{aligned}\hat{f}_n(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\hat{\phi}_X(t)}{\phi_\varepsilon(t)}\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi_K(th_n)\frac{\hat{\phi}_n(t)}{\phi_\varepsilon(t)}\,dt. \qquad (2.1)\end{aligned}$$

It is useful to note that the above estimator can be rewritten in the kernel form:

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_1^n K_n(\frac{x - X_j}{h_n}), \qquad (2.2)$$

where the kernel (called deconvoluted kernel) is given by

$$K_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\phi_K(t)}{\phi_\varepsilon(t/h_n)}\,dt. \qquad (2.3)$$

This deconvoluted kernel density estimator was considered by Stefanski and Carroll (1990) and Zhang (1990). Optimal properties are established by Fan (1991a). See also references given therein.

2.2 Deconvoluted kernel regression function estimators

Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the distribution of (X, Y), we first review the kernel estimator of the regression function $m(x) = E(Y|X = x)$. As before, let $K(\cdot)$ denote a kernel function and h_n be a bandwidth. Then the kernel estimator is the weighted average given by (Nadaraya (1964), Watson (1964))

$$\hat{m}_n(x) = \frac{\sum_j Y_j K\left(\frac{x-X_j}{h_n}\right)}{\sum_i K\left(\frac{x-X_i}{h_n}\right)}. \qquad (2.4)$$

Now suppose $X = X^\circ + \varepsilon$ and consider the problem of estimating the regression function $E(Y|X^\circ)$. Since the deconvoluted kernel (2.3) has the effect of separating X° from X in constructing the correct neighborhood around a given x, hence the kernel estimator (2.4) should be modified by replacing $K(\cdot)$ with $K_n(\cdot)$ to account for this effect. Based on this idea, Fan and Truong (1990) proposed the following estimator for $E(Y|X^\circ = x)$:

$$\hat{m}_n(x) = \frac{\sum_j Y_j K_n\left(\frac{x-X_j}{h_n}\right)}{\sum_i K_n\left(\frac{x-X_i}{h_n}\right)}. \qquad (2.5)$$

This estimator has many interesting optimal properties depending on the error distributions, which will be discussed with more details in Section 3.

2.3 Deconvoluted kernel median function estimators

Now consider the problem of estimating the conditional median function $m(x) = \mathrm{med}(Y|X^\circ)$. We first consider the problem of estimating the conditional distribution in the absence of measurement errors. A kernel estimator of $F(y|X = x) = P(Y \le y|X = x)$ is given by

$$F_n(y|x) = \frac{\sum_j K\left(\frac{x-X_j}{h_n}\right) 1(Y_j \le y)}{\sum_i K\left(\frac{x-X_i}{h_n}\right)}. \tag{2.6}$$

From this, one obtains the following conditional median estimator [see Truong (1989)]:

$$m_n(x) = F_n^{-1}(1/2|x). \tag{2.7}$$

To deal with the case involving measurement errors, we simply replace the kernel function $K(\cdot)$ by $K_n(\cdot)$ given in (2.3). This yields the deconvoluted kernel estimator of $F(y|X^\circ = x)$:

$$\hat{F}_n(y|x) = \frac{\sum_j K_n\left(\frac{x-X_j}{h_n}\right) 1(Y_j \le y)}{\sum_i K_n\left(\frac{x-X_i}{h_n}\right)}. \tag{2.8}$$

To estimate $m(x) = \mathrm{med}(Y|X^\circ = x)$, we propose

$$\hat{m}_n(x) = \hat{F}_n^{-1}(1/2|x). \tag{2.9}$$

More generally, it is easy to modify this approach to cover the problem of estimating conditional quantiles. Let $0 < p < 1$. The pth conditional quantile is defined as $F^{-1}(p|X^\circ = x)$, which is the pth quantile of the conditional distribution $F(\cdot|X^\circ = x)$. To estimate this quantile, we propose the following deconvoluted kernel estimator:

$$\hat{F}_n^{-1}(p). \tag{2.10}$$

Since (2.5) and (2.9) are respectively the mean and the median of (2.8), it is expected that the conditional median estimator (2.9) shares the similar optimalities as the regression estimator (2.5).

3 Optimal rates of convergence

Theoretical aspects of the problem on estimation of the density function f_{X° and the regression function $E(Y|X^\circ)$ are given in Fan (1991a, b) and Fan and Truong (1990). These results can be highlighted as follows.

- The estimates (2.1) and (2.5) are optimal in terms of rate of convergence.
- The rates of convergence depend on the smoothness of error distributions, which can be characterized into two categories: ordinary smooth and super smooth. Let ϕ_ε be the characteristic function of the error distribution. The distribution of ε is said to be
  - super smooth of order β: if the function $\phi_\varepsilon(\cdot)$ 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 function $\phi_\varepsilon(\cdot)$ 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, super smooth distributions are

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

  Ordinary smooth distributions are

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

  The rates of convergence for deconvoluted kernel estimators depend on β, the order of smoothness of the error distribution. They also depend on the smoothness of the regression function $m(x)$ and the marginal density function. For regression functions with bounded k-th derivatives, the following table illustrates the optimal rates of convergence according to the error distribution.

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

For density estimation based on deconvolution, a similar table is provided in Fan (1991a).

- The optimal choice of bandwidth h_n depends also on the error distribution. For the supersmooth error distribution of order β, the optimal $h_n = c(\log n)^{-1/\beta}$ for some constant c depending only on the error distribution and the kernel function. In the ordinary case, the optimal choice of bandwidth is $h_n = dn^{-\frac{1}{2k+2\beta+1}}$, for some constant d.

- Suppose $\min_{x\in[a,b]} f_{X^\circ}(x) > 0$ and that the conditional variance

$$\sigma^2(x) = \text{var}(Y|X^\circ = x) = \sigma^2.$$

Set

$$\hat{\sigma}_n^2 = \sum_1^n (Y_j - \hat{m}_n(X_j))^2 1_{\{a\le X_j\le b\}} / \sum_1^n 1_{\{a\le X_j\le b\}}$$

and

$$\hat{V}_n^2 = \frac{\hat{\sigma}_n^2}{n} \sum_1^n \frac{1}{h_n} K_n^2\left(\frac{x - X_j}{h_n}\right).$$

Then under appropriate conditions on h_n (which depends on the error distribution),

$$\sqrt{n}\frac{\hat{f}_n(x)[\hat{m}_n(x) - m(x)]}{\hat{V}_n} \longrightarrow N(0,1).$$

Note that this result is presented so that confidence intervals about the regression function can be easily obtained. For details, see Fan, Truong and Wang (1990). See also therein for a discussion on how to choose h_n for different type of error distributions. A similar result for density estimation is given Fan (1991b).

4 Extensions: a design-adaptive approach

4.1 Regression functions

As in the case of ordinary kernel regression function estimation, the optimal properties of the deconvoluted kernel regression estimator (2.5) depends on some smoothness condition of the marginal distribution. For example, to show that (2.5) possesses the optimal rates of convergence when the regression function $m(x)$ has a bounded $(k+1)$st derivative, it is also required that the marginal density function has a bounded $(k+1)$st derivative, too. See Fan and Truong (1990). In this section, we propose a different approach that would remove the extra smoothness condition imposed on the marginal distribution.

Note that the deconvoluted kernel regression estimator (2.5) can be viewed as $\hat{m}_n(x) = \hat{a}$ where $\hat{a}$ minimizes L^2-discrepancy:

$$G(a) = \sum_j (Y_j - a)^2 K_n\left(\frac{x - X_j}{h_n}\right),$$

with $K_n(\cdot)$ being the deconvoluted kernel function (2.3). Suppose now the regression function $m(x)$ has a bounded $(k+1)$st derivative. Define a new estimator by $\hat{m}_n(x) = \hat{a}_0$, where $\hat{a}_0, \ldots, \hat{a}_k$ minimize

$$G(a_0, \ldots, a_k) = \sum_j (Y_j - a_0 - a_1(x - X_j) - \cdots - a_k(x - X_j)^k)^2 K_n\left(\frac{x - X_j}{h_n}\right). \tag{4.1}$$

In fact, in the absence of measurement errors, Stone (1982) established that the estimators defined by (4.1) achieve the optimal rates of convergence without putting smoothness conditions on the marginal density, and Fan (1990) extended this idea to the smooth kernel case which possesses a number of efficient properties. The present approach is inspired by the latter idea based on smooth kernel and it is called design-adaptive since it does not require extra smoothness conditions on the marginal density function for achieving the optimal rates of convergence. In other words, the estimator adapts to both nearly uniform designs and highly clustered designs.

Example 1: **Weighted Linear Regression**

Suppose the regression function $m(x) = E(Y|X^\circ = x)$ has a bounded second derivative. Define $\hat{m}_n(x) = \hat{a}$, where $\hat{a}$, along with $\hat{b}$, minimize

$$G(a,b) = \sum_j (Y_j - a - b(x - X_j))^2 K_n\left(\frac{x - X_j}{h_n}\right).$$

Simple algebra shows

$$\hat{m}_n(x) = \sum_1^n w_{n,j} Y_j / \sum_1^n w_{n,i},$$

where

$$w_{n,j} = K_n\left(\frac{x - X_j}{h_n}\right)(s_{n,2} - (x - X_j)s_{n,1})$$

and

$$s_{n,l} = \sum_{j=1}^n K_n\left(\frac{x - X_j}{h_n}\right)(x - X_j)^l, \qquad l = 1, 2.$$

4.2 Quantile functions

The deconvoluted kernel median function estimator (2.9) can also be viewed as $\hat{m}_n(x) = \hat{a}$ where $\hat{a}$ minimizes L^1-discrepancy:

$$G(a) = \sum_j |Y_j - a| K_n\left(\frac{x - X_j}{h_n}\right),$$

with $K_n(\cdot)$ given by (2.3). Similar to the approach given in the previous section, define a new estimator by $\hat{m}_n(x) = \hat{a}_0$, where $\hat{a}_0, \ldots, \hat{a}_k$ minimize

$$G(a_0, \ldots, a_k) = \sum_j |Y_j - a_0 - a_1(x - X_j) - \cdots - a_k(x - X_j)^k| K_n\left(\frac{x - X_j}{h_n}\right). \tag{4.2}$$

The estimator defined by (4.2) can further be extended as follows. Let $\hat{m}_n(x) = \hat{a}_0$, with $\hat{a}_0, \ldots, \hat{a}_k$ minimizing

$$G(a_0, \ldots, a_k) = \sum_j g(Y_j - a_0 - a_1(x - X_j) - \cdots - a_k(x - X_j)^k) K_n\left(\frac{x - X_j}{h_n}\right), \tag{4.3}$$

where $g(z) = (pz_+ + (1-p)z_-)$, $0 < p < 1$. This leads to the estimation of conditional quantiles in general. Note that, when $p = 1/2$, (4.3) reduces to (4.2).

5 Numerical Results

In this section, we are interested in the estimation of regression function for binary responses. Let Y denote a 0-1 random variable whose distribution depends on another random variable (covariate) X° so that the logistic transformation of its regression function is quadratic (see Figure 2):

$$m(x) = E(Y|X^\circ = x) = P(Y = 1|X^\circ = x) = \frac{\exp(-6(x-0.5)^2)}{1+\exp(-6(x-0.5)^2)}.$$

That is, given $X^\circ = x$, Y takes the value 1 with probability $m(x)$. Now suppose $X^\circ \sim \text{unif}(0,1)$ and that X° is not available, instead we observe it through $X = X^\circ + \varepsilon$, where ε is a random error with mean 0 and finite variance. Given a random sample from the distribution of (X, Y), we would like to estimate $m(x) = E(Y|X^\circ = x)$. From the theoretical aspect of Section 3, we know that the error distribution plays an important role in determining the sampling behaviors of the deconvoluted kernel estimator. Computationally, this will be addressed in the following examples.

Example 2: **Super Smooth Errors**

Suppose ε has a normal distribution with mean 0 and variance σ_0^2. Let $(X_1, Y_1), \ldots, (X_{200}, Y_{200})$ denote a random sample from the distribution of (X, Y) so that

$$\begin{aligned} X_i &= X_i^\circ + \varepsilon_i, \quad X_i^\circ \sim_{\text{iid}} \text{unif}(0,1), \quad \varepsilon_i \sim_{\text{iid}} N(0, \sigma_0^2), \\ Y_i &= \begin{cases} 1, & \text{with probability } m(X_i^\circ); \\ 0, & \text{with probability } 1 - m(X_i^\circ). \end{cases} \end{aligned} \tag{5.1}$$

Here σ_0 is selected so that $\sigma_0^2/\text{var}(X_i^\circ) = 0.10$. Suppose the kernel $K(\cdot)$ is an inverse triangular density so that its Fourier transform is given by

$$\phi_K(t) = (1 - |t|)_+.$$

By $\phi_\varepsilon(t) = \exp(-\frac{1}{2}\sigma_0^2 t^2)$ and (2.3),

$$K_n(x) = \frac{1}{\pi}\int_0^1 \cos(tx)(1-t)\exp\left(\frac{\sigma_0^2 t^2}{2h_n^2}\right) dt. \tag{5.2}$$

For the estimator (2.5) to achieve the optimal rates of convergence, the bandwidth h_n is chosen so that $h_n = c\sigma_0(\log n)^{-1/2}$ with $c > 1$. See Fan and Truong (1990). Note that the deconvoluted kernel function depends on n. Figure 1 plots these functions for different values of the constant factor c. Deconvoluted estimators with different c are presented in Figure 2.

Deconvoluted kernel and ordinary kernel functions

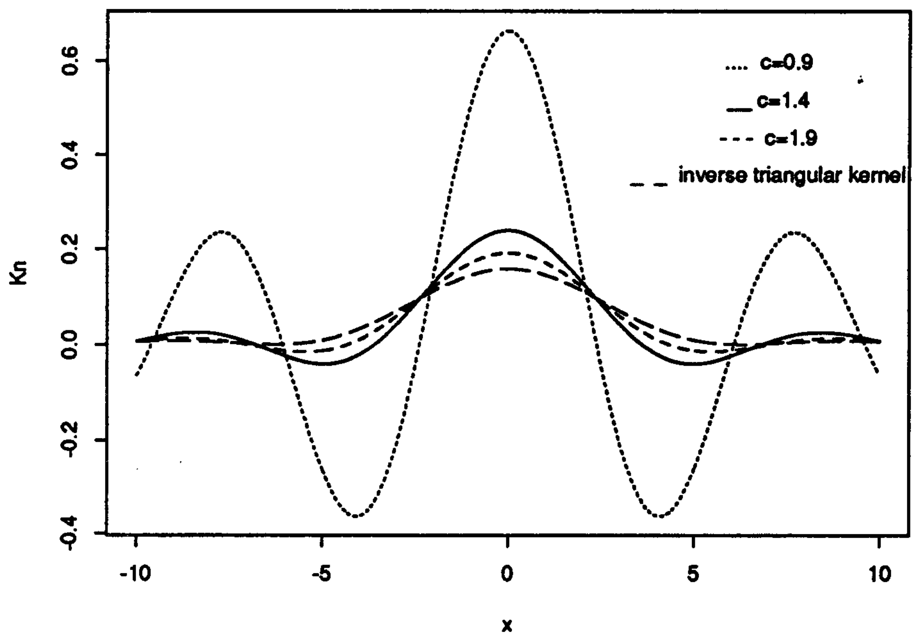

Figure 1

Deconvoluted kernel estimates

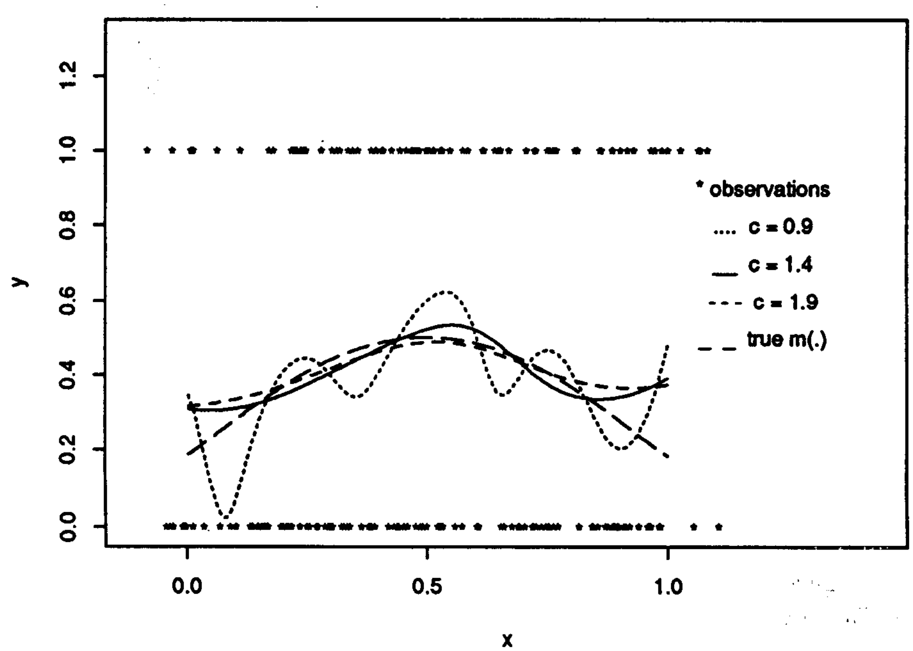

Figure 2

Example 3: **Ordinary Smooth Errors**

As in (5.1), $(X_1, Y_1), \ldots, (X_{200}, Y_{200})$ denotes a random sample from the distribution of (X, Y) with ε now having a double exponential distribution:

$$f_\varepsilon(z) = \sigma_0^{-1} \exp(-2|z|/\sigma_0).$$

Here σ_0 is chosen so that $\sigma_0^2/\mathrm{var}(X_i^\circ) = 0.10$. The 200 simulated data are plotted in Figure 4. A star " * " indicates an observed data point.

Note that

$$\phi_\varepsilon(t) = \frac{1}{1 + \frac{1}{4}\sigma_0^2 t^2}.$$

If $K(\cdot)$ is the gaussian kernel $K(x) = (\sqrt{2\pi})^{-1} \exp(-x^2/2)$, then by (2.3),

$$\begin{aligned} K_n(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_K(t) \left(1 + \frac{\sigma_0^2 t^2}{4h_n^2}\right) dt \\ &= K(x) + \frac{\sigma_0^2}{4h_n^2} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) t^2 \phi_K(t)\, dt \\ &= K(x) - \frac{\sigma_0^2}{4h_n^2} K''(x) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \left[1 - \frac{\sigma_0^2}{4h_n^2}(x^2 - 1)\right]. \end{aligned}$$

Figure 3 plots the deconvoluted kernel functions and Figure 4 gives the estimators (2.5) with $h_n = 0.08, 0.16, 0.32$. It is clear that the estimate with $h_n = 0.08$ is under smooth the curve, while the estimate with $h_n = 0.32$ over smooth the curve.

Fan and Truong (1990) showed that the optimal rates of convergence can be achieved by choosing $h_n \sim cn^{-1/9}$ with $c > 0$.

We compute also the estimate (2.5) with other choices of bandwidth. It turns out that the estimates for the super smooth case are very sensitive to the choice of the bandwidth when $h_n = c\sigma_0(\log n)^{-1/2}$ with $c < 1$. This seems compatible with the theory: the variance is very large and the estimate is even not *consistent*. One the other hand, the bandwidth is relatively less sensitive for the ordinary smooth case.

Deconvoluted kernel and ordinary kernel functions

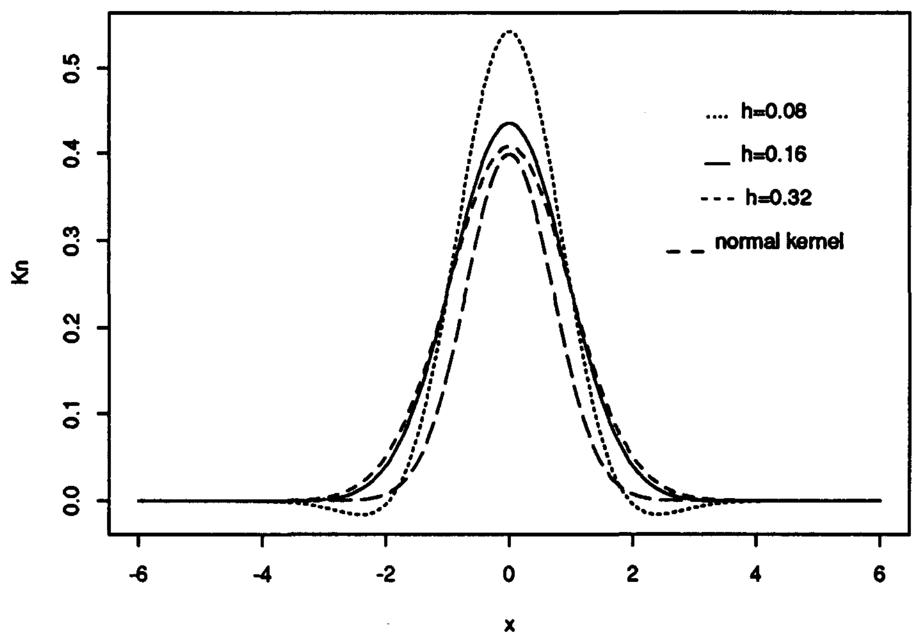

Figure 3

Deconvoluted kernel estimates

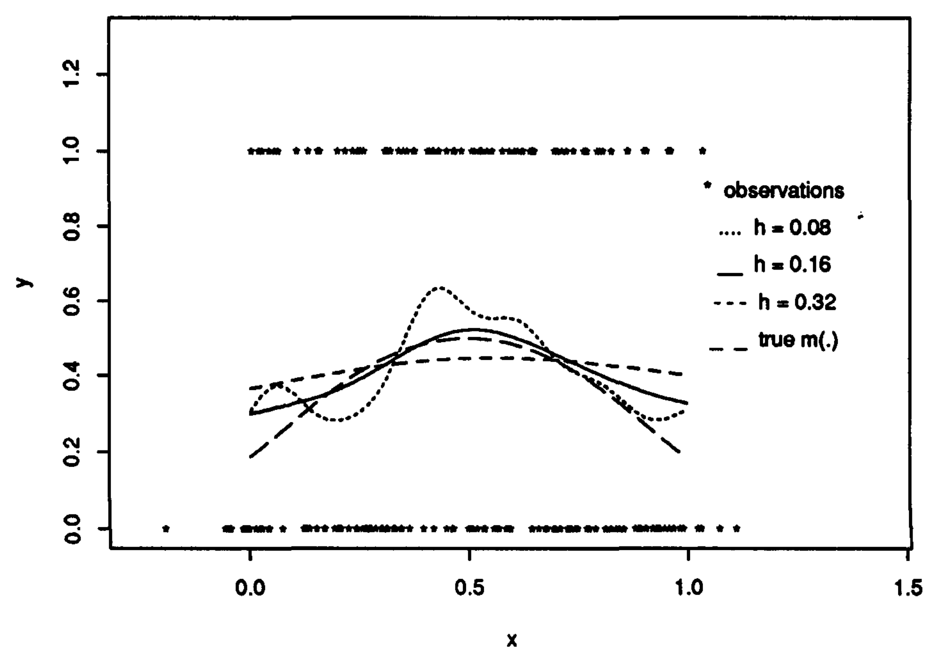

Figure 4

6 Conclusions

In this paper, we discuss methods for estimating nonparametric functions involving errors-in-variables. Density estimation was considered first because it sets the foundation for *deconvolution*. We then address problems on estimating the regression function and conditional quantiles. Extensions of these estimators to design-adaptive ones (i.e. local polynomials) are also given. Although we have not provided proofs, but it seems plausible that the design-adaptive estimators would achieve the optimal rates of convergence without requiring extra smoothness conditions on the marginal distribution. There are also some other important open problems:

- the performance of the deconvoluted conditional quantile estimators (2.7) and (2.10);
- the sampling properties of design-adaptive quantile estimators (4.2) and (4.3);
- the amount of noise level $\sqrt{\mathrm{var}(\varepsilon)}$ that would make nonparametric deconvolution feasible.

References

[1] Armstrong, B. (1985). Measurement error in the generalized linear models. *Communications in Statistics — Simulation and Computation*, **14**, 529–544.

[2] Fan, J. (1990). A remedy to regression estimators and nonparametric minimax efficiency. *Institute of Statistics Memio Series #2028*, University of North Carolina, Chapel Hill.

[3] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problem. To appear in *Ann. Statist.*

[4] Fan, J. (1991b). Asymptotic normality for deconvolving kernel density estimators. To appear in *Sankhyā.*

[5] Fan, J. and Truong, Y. (1990). Nonparametric regression with errors-in-variables. *Institute of Statistics Mimeo Series #2023*, Univ. of North Carolina, Chapel Hill.

[6] Fan, J., Truong, Y. and Wang, Y. (1990). Confidence intervals for non-parametric regression with errors-in-variables. *Manuscript.*

[7] Fuller, W. A. (1987). *Measurement error models.* Wiley, New York.

[8] Nadaraya, E. A. (1964). On estimating regression. *Theor. Probability Appl.*, **9**, 141–142.

[9] 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.

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

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

[12] Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, **21**, 169-184.

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

[14] Truong, Y.K. (1989). Asymptotic properties of kernel estimators based on local medians. *Ann. Statist.*, **17**, 606-617.

[15] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, **26**, 359–372.

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

[17] Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, **18**, 806–830.