

# ASYMPTOTIC NORMALITY OF PARAMETRIC PART IN PARTIALLY LINEAR MODELS WITH MEASUREMENT ERROR IN THE NONPARAMETRIC PART \*

Hua Liang

Institut für Statistik und Ökonometrie  
Humboldt-Universität zu Berlin  
D-10178 Berlin, Germany

## Abstract

We consider the partially linear model relating a response  $Y$  to predictors  $(X, T)$  with mean function  $X^T\beta + g(T)$  when the  $T$ 's are measured with additive error. We derive an estimator of  $\beta$  by modification local-likelihood method. The resulting estimator of  $\beta$  is shown to be asymptotically normal.

**Key Words and Phrases:** Errors-in-Variables; Measurement Error; Nonparametric Likelihood; Partially Linear Model; Semiparametric Models.

**Short title:** Partially Linear Measurement Error Models.

**AMS 1991 subject classification:** Primary: 62J99, 62H12, 62E25, 62F10 Secondary: 62H25, 62F10, 62F12, 60F05.

## 1 INTRODUCTION AND BACKGROUND

The interest in study measurement error model is growing with the publication of a series of papers on various topics, especially the monograph of Carroll, et al. (1995). Broadly it can be divided into two parts: the first one focuses to linear measurement error model. See Anderson (1984), Carroll, et al. (1984) and Stefanski (1985). The second one mainly deals with nonlinear measurement error models. See Fan and Truong (1993) and references therein. Liang, Härdle and Carroll (1997) firstly considered the semiparametric partially linear model relating a response  $Y$

---

\*The author would like to thank Professor Raymond J. Carroll and Mr. Sommerfeld Volker for their constructive suggestions and valuable comments which greatly improved the presentation of this paper. This research was supported by Alexander von Humboldt Foundation and in part by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse".

to predictors  $(X, T)$  with function  $X^T\beta + g(T)$  when the  $X$ 's are unobserved and with additive error. The authors derived an estimator of  $\beta$ , which is shown to be consistent and asymptotically normal. If we interchange the roles of  $X$  and  $T$ , so that the parametric part is measured exactly and nonparametric part is measured with error,  $E(Y|X, T) = \theta X + g(T)$  and  $W = T + U$ , where  $U$  is measurement error. How about the result in this situation? Liang, Härdle and Carroll. (1997) conjectured that  $\theta$  is estimable at parametric rates. The goal of this paper is just to present a detailed and positive answer.

Fan and Truong (1993) have treated the case that  $\beta = 0$  and  $T$  has measurement error. They have proved that the nonparametric function  $g(\cdot)$  can be estimated only at logarithmic rates, but not with rate  $n^{-2/5}$  as usual even with normally distributed measurement error.

Consider the semiparametric partially linear model based on a sample of size  $n$ ,

$$Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, \quad (1)$$

where  $X_i$  is a random vector,  $T_i$  is a random variable defined on  $[0, 1]$ , the function  $g(\cdot)$  is unknown, and the model errors  $\varepsilon_i$  are independent with conditional mean zero given the covariates. In this model, the covariates  $T$  are measured with error, and instead of observing  $T$ , we observe

$$W_i = T_i + U_i, \quad (2)$$

where the measurement errors  $U_i$  are independent and identically distributed, independent of  $(Y_i, X_i, T_i)$ , with mean zero and covariance matrix  $\Sigma_{uu}$ . We will assume that  $U$  has a known distribution, which is proposed by Fan and Truong (1993) in order to assure that the model is identifiable.

When  $T$ 's are observable, in literature authors often constructed root- $n$  consistent estimator of  $\beta$  by local-likelihood algorithm (See Engle, et al. (1986), Heckman (1986), Chen (1988), Speckman (1988), Cuzick (1992a,b) and Severini & Staniswalis (1994)) as follows. To fix the parametric component  $\beta$  and obtain an estimate  $\hat{g}(T, \beta)$  of the nonparametric component  $g(\cdot)$  using some kind of smoothing method. For example, in the Severini and Staniswalis implementation,  $\hat{g}(T, \beta)$  maximizes a weighted likelihood assuming that the model errors  $\varepsilon_i$  are homoscedastic and normally distributed, with the weights being kernel weights with symmetric kernel density function  $K(\cdot)$  and bandwidth  $h$ .  $\hat{g}(T, \beta)$  is then used to obtain an estimator of the parametric component of model (1), using methods such as maximum likelihood or least squares. For example let the solution of

$$\text{minimize } \sum_{i=1}^n \left\{ Y_i - X_i^T\beta - \hat{g}(T_i, \beta) \right\}^2 \quad (3)$$

as the estimate for  $\beta$ , which can be determined explicitly by a projected least squares algorithm. Let  $\hat{g}_{y,h}(\cdot)$  and  $\hat{g}_{x,h}(\cdot)$  be the kernel regressions with bandwidth  $h$  of  $Y$  and  $X$  on  $T$ , respectively. Then

$$\beta_n = \left[ \sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{X_i - \hat{g}_{x,h}(T_i)\}^T \right]^{-1} \sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{Y_i - \hat{g}_{y,h}(T_i)\}. \quad (4)$$

It has shown that the estimator (4) does not require undersmoothing and the usual bandwidth with order  $h \sim n^{-1/5}$  leads to that  $\beta_n$  is asymptotically normal with mean zero and variance  $B^{-1}CB^{-1}$ , where  $B$  is the covariance matrix of  $X - E(X|T)$  and  $C$  is the covariance matrix of  $\varepsilon\{X - E(X|T)\}$ .

Due to the disturbance of measurement error  $U$ , the least squares form of (4) has to be modified, otherwise  $\beta_n$ , more exactly  $\hat{g}_{x,h}(T_i)$  and  $\hat{g}_{y,h}(T_i)$ , are not be statistic any more. In next section, we will redefine an estimator of  $\beta$ . More exactly, we have to search a new estimator of  $g(\cdot)$ , and then the regression  $Y$  and  $X$  on  $W$ . Section 3 states our main result. Section 4 provides some simulation studies. Several remarks are given in Section 5. All technical proofs are postponed in Appendix.

## 2 CONSTRUCTION OF ESTIMATORS

As pointed out in the former section, our first obstacle is how to estimate the nonparametric function  $g(\cdot)$  under the  $T$  being unobserved. This can be overcome by borrowing the ideas of Fan and Truong (1993). That is, using deconvolution technology, one can construct consistent nonparametric estimates of  $g(\cdot)$ , which has convergence rate in some sense under appropriate assumptions.

First, we briefly describe deconvolution method, which has been studied by Stefanski and Carroll (1990), Fan and Truong (1993). Denote the densities of  $W$  and  $T$  by  $f_W(\cdot)$  and  $f_T(\cdot)$ , respectively. As pointed in literature,  $f_T(\cdot)$  can be estimated by using the estimator

$$\hat{f}_n(t) = \frac{1}{nh_n} \sum_{j=1}^n K_n \left( \frac{t - W_j}{h_n} \right)$$

with

$$K_n(t) = \frac{1}{2\pi} \int_{R^1} \exp(-ist) \frac{\phi_K(s)}{\phi_U(s/h_n)} ds \quad (5)$$

where  $\phi_K(\cdot)$  is the Fourier transform of  $K(\cdot)$ , a kernel function,  $\phi_U(\cdot)$  is the characteristic function of the error variable  $U$ . More detailed discussion see Fan and Truong (1993).

Denote

$$\omega_{ni}(\cdot) = K_n \left( \frac{\cdot - W_i}{h_n} \right) / \sum_j K_n \left( \frac{\cdot - W_j}{h_n} \right) \stackrel{\text{def}}{=} \frac{1}{nh_n} K_n \left( \frac{\cdot - W_i}{h_n} \right) / \hat{f}_n(\cdot)$$

Next let us return to our goal. Notice the fact  $g(t) = E(Y - X^T\beta|T = t)$ , as dealt by Fan and Truong (1993) one can define

$$g_n(t) = \sum_{i=1}^n \omega_{ni}(t)(Y_i - X_i^T\beta)$$

as the estimator of  $g(\cdot)$  if  $\beta$  were known.

Let  $\hat{\beta}_n$  be the solution (3) after in which  $\hat{g}(T_i, \beta)$  substituting by  $g_n(W_i)$ . Then the generalized least squares estimator  $\hat{\beta}_n$  of  $\beta$  can be explicitly indicated as

$$\hat{\beta}_n = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} (\tilde{\mathbf{X}}^T \tilde{\mathbf{Y}}) \quad (6)$$

where  $\tilde{\mathbf{Y}}$  denotes  $(\tilde{Y}_1, \dots, \tilde{Y}_n)$  with  $\tilde{Y}_i = Y_i - \sum_{j=1}^n \omega_{nj}(W_i)Y_j$  and  $\tilde{\mathbf{X}}$  denotes  $(\tilde{X}_1, \dots, \tilde{X}_n)$  with  $\tilde{X}_i = X_i - \sum_{j=1}^n \omega_{nj}(W_i)X_j$ .

The estimator  $\hat{\beta}_n$  will be shown to process asymptotic normality, which forms the core of this paper. Up to now, we solve the first problem. The further study is to develop asymptotic theory for the estimator  $\hat{\beta}_n$ .

### 3 MAIN RESULTS

We make the following assumptions. Firstly some notations are introduced.  $\gamma_j(t) = E(x_{ij}|T_i = t)$ ,  $V_{ij} = x_{ij} - \gamma_j(T_i)$  for  $i = 1, \dots, n$  and  $j = 1, \dots, p$ .

**Assumption 1.1.**  $\sup_{0 \leq t \leq 1} E(\|X_1\|^3|T = t) < \infty$  and  $E(V_1 V_1^T) = B$  and  $B$  is a positive definite matrix, where  $V_i = (V_{i1}, \dots, V_{ip})^T$ .

**Assumption 1.2.**  $g(\cdot)$  and  $\gamma_j(\cdot)$  are Lipschitz continuous of order 1.

**Assumption 1.3.** (i) The marginal density  $f_T(\cdot)$  of the unobserved is bounded away from 0 on  $[0, 1]$ , and has a bounded  $k$ -th derivative. Here and below  $k$  is a integer.

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

(iii) The distribution of the error  $\varepsilon$  is super smooth or ordinary smooth.

The definitions of super smooth and ordinary smooth distributions are given by Fan and Truong (1993). We also state them for easy reference.

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

$$d_0|t|^{\alpha_0} \exp(-|t|^\alpha/\zeta) \leq |\phi_\varepsilon(t)| \leq d_1|t|^{\alpha_1} \exp(-|t|^\alpha/\zeta) \quad \text{as } t \rightarrow \infty, \quad (7)$$

where  $d_0, d_1, \alpha$  and  $\zeta$  are positive constants and  $\alpha_0$  and  $\alpha_1$  constants.

2. Ordinary smooth of order  $\alpha$ : If the characteristic function of the error distribution  $\phi_\varepsilon(\cdot)$  satisfies

$$d_0|t|^{-\alpha} \leq |\phi_\varepsilon(t)| \leq d_1|t|^{-\alpha} \quad \text{as } t \rightarrow \infty, \quad (8)$$

for positive constants  $d_0, d_1, \alpha$ .

For example, standard normal and Cauchy distributions are super smooth with  $\alpha = 2$  and  $\alpha = 1$  respectively; Gamma distribution of degree  $p$  and double exponential distribution are ordinary smooth ones with  $\alpha = p$  and  $\alpha = 2$  respectively.

**Assumption 1.4.** *The kernel  $K(\cdot)$  is a  $k$ -th order kernel function, that is*

$$\int_{-\infty}^{\infty} K(u)du = 1, \quad \int_{-\infty}^{\infty} u^l K(u)du \begin{cases} = 0 & l = 1, \dots, k-1 \\ \neq 0 & l = k \end{cases}$$

Assumptions 1.1 and 1.2 are required to establish asymptotic normality with observed values  $T_i$ . Assumptions 1.3 and 1.4 ensure that our estimator of nonparametric function  $g(\cdot)$  are still satisfied with the analogous conclusion given by Fan and Truong (1993).

Our main result concerns the limit distributions of the estimate of  $\beta$ , which is stated as follows.

**THEOREM.** *Suppose Assumptions 1.1-1.4 hold and  $E(\varepsilon^3 + \|U\|^3) < \infty$ . If either of the follow two items is proceeding, that is,*

(i) *When we have super smooth error distribution.  $\phi_K(t)$  has a bounded support on  $|t| \leq M_0$  and we take the bandwidth  $h_n = c(\log n)^{-1/\alpha}$  with  $c > M_0(2/\zeta)^{1/\alpha}$ ;*

(ii) *When we have ordinary smooth error distribution.*

$$t^\alpha \phi_\varepsilon(t) \rightarrow c, \quad t^{\alpha+1} \phi'_\varepsilon(t) = O(1) \quad \text{as } t \rightarrow \infty$$

for some constant  $c \neq 0$ , and

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

We take  $h_n = dn^{-1/(2k+2\alpha+1)}$  with  $d > 0$ .

Then  $\hat{\beta}_n$  is an asymptotically normal estimator, i.e.

$$n^{1/2}(\hat{\beta}_n - \beta) \Rightarrow N(0, \sigma^2 B^{-1})$$

where  $B$  is given in Assumption 1.1.

## 4 SIMULATION

We performed a moderate sample Monte-Carlo simulation to show the behaviour of the estimator  $\hat{\beta}_n$ . A generalization of the model studied by Fan and Truong (1993) is considered.

$$Y = X^T \beta + g(T) + \varepsilon \quad \text{and} \quad W = T + U \quad \text{with} \quad \beta = 0.75$$

Here  $X \sim N(0, 1)$ ,  $T \sim N(0.5, 0.25^2)$ ,  $\varepsilon \sim N(0, 0.0015^2)$ ,  $U \sim N(0, 0.125^2)$ ,  $g(t) = t_+^3(1-t)_+^3$ .

Suppose the function  $K(\cdot)$  has a Fourier transform by  $\phi_K(t) = (1-t^2)_+^2$ . By (5),

$$K_n(t) = \frac{1}{\pi} \int_0^1 \cos(st)(1-s^2)^3 \exp\left(\frac{0.125^2 s^2}{2h_n^2}\right) ds \quad (9)$$

For the above model, we use two different kernels: (9) and quartic kernel  $(15/16)(1-u^2)^2 I(|u| \leq 1)$  (ignoring measurement error). Our aim is to compare the results in cases of considering measurement error and ignoring measurement errors. The results with different sample-numbers will be presented. The simulation number is  $N = 100$ .

Each simulation result is taken as an sample. Basing these  $N$  samples, its sample moment is taken as our estimate value. The mean square error (MSE) is also calculated. Table 1 gives the final detailed simulation results.

TABEL 1: Simulation results ( $\times 10^{-3}$ )

Kernel	$n = 100$		$n = 500$		$n = 1000$		$n = 2000$	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(9)	2.249	0.444	2.247	0.351	2.288	0.233	2.234	0.153
quartic	2.272	0.527	2.256	0.361	2.251	0.274	2.237	0.166

In simulation procedure, we also arrive at the nonparametric fitting, that is,

$$\hat{g}_n(t) = \sum_{i=1}^n \omega_{ni}(t)(Y_i - X_i^T \hat{\beta}_n) \quad (10)$$

and  $\hat{\beta}_n$  is the resulting estimator given in (6).

An analysis ignoring measurement error (with quartic kernel) found some curvature in  $T$ , see Figure 1 for the comparison of  $g(T)$  with its estimator (10) using the different-size samples. The bandwidth used in our simulation is selected using crossvalidation to predict the response. More precisely, we compute the average squared error using a geometric sequence of 41 bandwidths ranging in  $[0.1, 0.5]$ . The optimal bandwidth is selected to minimize the average squared error

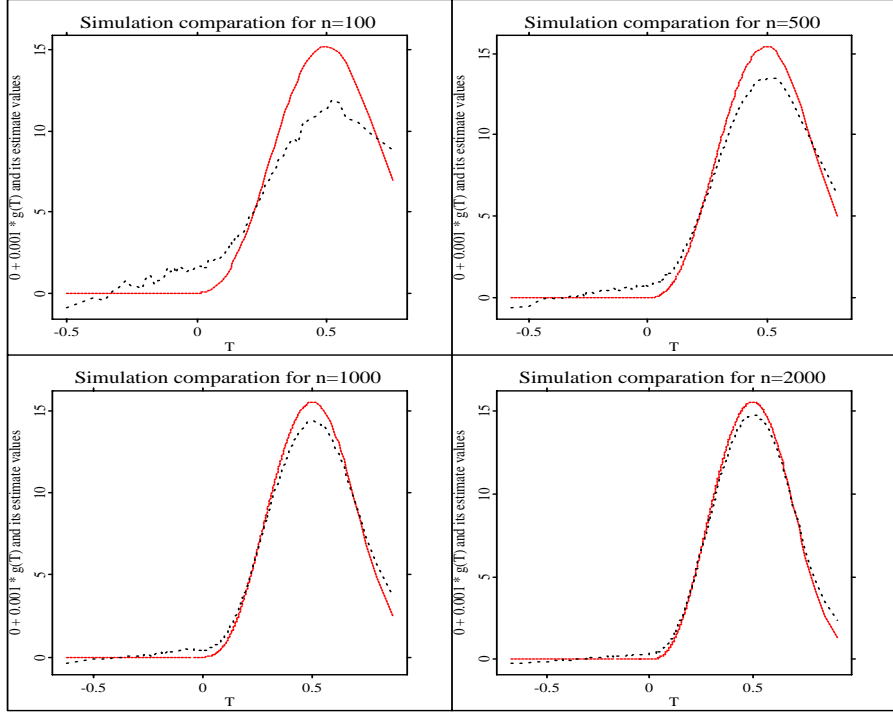


Figure 1: Estimates of the function  $g(T)$

among 41 candidates. The solid lines stand for true values and the dashed lines do the value of the resulting estimator given by (10). Table 1 and the four pictures in Figure 1 illustrate that our estimators for parametric and nonparametric parts are both satisfied.

## 5 DISCUSSION

The theorem shows us that in large sample there is no cost due to measurement error of  $T$ . That is, the estimator of  $\beta$  given by (6) is equivalent to the estimator given by (4) when we suppose that  $T_i$  were known. This phenomena looks like unreasonable until one observe that measurement error of  $T$  mainly affect the estimator of the nonparametric part so that the estimator of the nonparametric part can not reach nonparametric-rate as usual. In addition,  $k$  and  $\alpha$  do not affect our asymptotic normal result as long  $\log^{-k/\alpha} n$  converges to zero and the Lemma A.2 holds. On the other hand, the estimator of nonparametric part which reaches logarithmic rates, decided by  $k$  and  $\alpha$  and then  $\log^{-k/\alpha} n$ , is easily derived by directly copying the related procedure of Fan and Truong (1993).

The cases that the parametric part  $X$  of the model has measurement error and the nonparametric part  $T$  is measured exactly or that the parametric part is measured exactly and the non-

parametric part is measured with error have been discussed. In more general situation, that is  $X$  and  $T$  are both observed with measurement errors. How to seeking parametric rate estimator of parameter  $\beta$  is an interesting issue, which is still open one.

## 6 APPENDIX

Lemma A.1 gives a rather general result on strong uniform convergence, which is applied in the proofs of the present context. Its proof is referred to Liang and Härdle (1997).

**Lemma A.1.** *Let  $V_1, \dots, V_n$  be independent random variables with 0 means and  $\sup_j E|V_j|^r \leq C < \infty$  ( $r \geq 2$ ). Assume  $\{a_{ki}, k, i = 1, \dots, n\}$  be a sequence of positive numbers such that  $\sup_{i,k \leq n} |a_{ki}| \leq n^{-p_1}$  for some  $0 < p_1 < 1$  and  $\sum_{j=1}^n a_{ji} = O(n^{p_2})$  for  $p_2 \geq \max(0, 2/r - p_1)$ . Then*

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n a_{ki} V_k \right| = O(n^{-s} \log n) \quad s = (p_1 - p_2)/2. \quad a.s.$$

Lemma A.2 provides bounds for  $\gamma_j(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) \gamma_j(T_k)$  and  $g(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) g(T_k)$ . The proof is mainly based upon the conclusion of Fan and Truong (1993).

**Lemma A.2.** Suppose that Assumptions 1.1 and 1.4 hold. Then

$$\max_{1 \leq i \leq n} \left| G_j(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) G_j(T_k) \right| = o(1) \quad \text{for } j = 0, \dots, p,$$

where  $G_0(\cdot) = g(\cdot)$  and  $G_l(\cdot) = \gamma_l(\cdot)$  for  $l = 1, \dots, p$ .

Theoretically, if  $\beta$  were known, absorbing  $X^T \beta$  into  $Y$ ,  $\sum_{k=1}^n \omega_{nk}(w)(Y_k - X_k^T \beta)$  is parallel to the estimate of nonparametric function proposed by Fan and Truong (1993). No matter what the error distribution is, one can find from the main results of Fan and Truong (1993) that

$$\max_{1 \leq i \leq n} \left| g(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) \{g(T_k) + \varepsilon_k\} \right| = O(\log^{-k/\alpha} n)$$

It can be derived that

$$\sup_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(W_i) \varepsilon_k \right| = o(1)$$

by taking  $V_i = \varepsilon_i$  and  $a_{ki} = \omega_{nk}(W_i)$  in Lemma A.1. These arguments imply that

$$\max_{1 \leq i \leq n} \left| g(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) g(T_k) \right| = o(1)$$

The proofs for  $\gamma_l(\cdot)$  ( $l = 1, \dots, p$ ) are similarly as the proof of Lemma 2 of Fan and Truong (1993). More precisely, noting the Lipschitz continuity of  $\gamma_l(\cdot)$  one obtains that

$$\max_{1 \leq i \leq n} \left| \gamma_l(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) \gamma_l(T_k) I(|T_i - T_k| \leq c_n) \right| = O(c_n)$$



On the other hand

$$\sum_{k=1}^n \omega_{nk}(W_i) I(|T_i - T_k| > c_n) = \sum_{k=1}^n \frac{K_n\left(\frac{W_i - W_k}{h_n}\right) I(|T_i - T_k| > c_n)}{\sum_{j=1}^n K_n\left(\frac{W_i - W_j}{h_n}\right)} \quad (11)$$

Adopting the proof of Lemma 2 of Fan and Truong (1993), the orders of the denominator and numerator of (11) can be shown to be equal to the orders of  $nc_n h_n$  and  $nh_n$ , respectively. Letting  $c_n$  tend to zero, the fact that Lemma A.2 hold for  $\gamma_l(\cdot)$  is immediately derived.

**Lemma A.3.** If Assumptions 1.1-1.4 hold. Then

$$\lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = B$$

**Proof.** Denote  $\bar{\gamma}_{ns}(T_i) = \gamma_s(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) X_{ks}$ . It follows from  $X_{js} = \gamma_s(T_j) + V_{js}$  that the  $(s, m)$  element of  $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$  ( $s, m = 1, \dots, p$ ) is

$$\begin{aligned} \sum_{j=1}^n \tilde{X}_{js} \tilde{X}_{jm} &= \sum_{j=1}^n V_{js} V_{jm} + \sum_{j=1}^n \bar{\gamma}_{ns}(T_j) V_{jm} + \sum_{j=1}^n \bar{\gamma}_{nm}(T_j) V_{js} + \sum_{j=1}^n \bar{\gamma}_{ns}(T_j) \bar{\gamma}_{nm}(T_j) \\ &\stackrel{\text{def}}{=} \sum_{j=1}^n V_{js} V_{jm} + \sum_{q=1}^3 R_{nsm}^{(q)} \end{aligned}$$

The strong law of large number implies that  $\lim_{n \rightarrow \infty} 1/n \sum_{i=1}^n V_i V_i^T = B$ . Notice that  $\bar{\gamma}_{ns}(T_i)$  is just  $\tilde{\gamma}_s(T_i) + \sum_{k=1}^n \omega_{nk}(W_i) V_{ks}$ . In Lemma A.1, taking  $a_{ik} = \omega_{nk}(W_i)$  and  $V_k = V_{ks}$  and  $p_1 = 2, p_2 = 0$  one obtains that  $\sum_{k=1}^n \omega_{nk}(W_i) V_{ks} = o(1)$ , which and Lemma A.2 lead  $\bar{\gamma}_{ns}(T_i) = o(1)$ . This means  $R_{nsm}^{(3)} = o(n)$ , which together with the Cauchy-Schwarz inequality show that  $R_{nsm}^{(1)} = o(n)$  and  $R_{nsm}^{(2)} = o(n)$ . This completes the proof of the lemma.

**PROOF OF THE THEOREM.** We firstly outline the proof of the theorem. We decompose  $\sqrt{n}(\hat{\beta}_n - \beta)$  into three terms. Then we will calculate the tail probability value of each term. By the definition of  $\hat{\beta}_n$ ,

$$\begin{aligned} \sqrt{n}(\hat{\beta}_n - \beta) &= \sqrt{n}(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \left[ \sum_{i=1}^n \tilde{X}_i g_{ni} - \sum_{i=1}^n \tilde{X}_i \left\{ \sum_{j=1}^n \omega_{nj}(W_i) \varepsilon_j \right\} + \sum_{i=1}^n \tilde{X}_i \varepsilon_i \right] \\ &\stackrel{\text{def}}{=} A(n) \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i g_{ni} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i \left\{ \sum_{j=1}^n \omega_{nj}(W_i) \varepsilon_j \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i \varepsilon_i \right] \quad (12) \end{aligned}$$

where  $A(n) = n^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$  and  $g_{ni} = g(T_i) - \sum_{k=1}^n \omega_{nk}(W_i) g(T_k)$ .

Lemma A.3 means that  $A(n)$  converges to  $B^{-1}$ . Thus our problem is to prove the first and the second terms in the parentheses of the right-hand side of (12) converge in probability to zero and  $1/\sqrt{n} \sum_{i=1}^n \tilde{X}_i \varepsilon_i$  converges to normal distribution with mean zero and covariate matrix  $\sigma^2 B$ . The latter half assertion can be shown by using central limit theorem and Lemma A.3, or referring Chen (1988). Let us now verify the former assertion.

Taking  $r = 3$ ,  $V_k = \varepsilon_k$  or  $V_{kl}$ ,  $a_{ji} = \omega_{nj}(W_i)$ ,  $p_1 = 2/3$  and  $p_2 = 0$  in Lemma A.1, one obtains the following equations, which will play critical roles in the processes of the proofs.

$$\max_{i \leq n} \left| \sum_{k=1}^n \omega_{nk}(W_i) \varepsilon_k \right| = O(n^{-1/3} \log n) \quad a.s. \quad (13)$$

$$\max_{i \leq n} \left| \sum_{k=1}^n \omega_{nk}(W_i) V_{kl} \right| = O(n^{-1/3} \log n) \quad \text{for } l = 1, \dots, p \quad a.s. \quad (14)$$

Notice that

$$\sum_{i=1}^n \tilde{x}_{ij} g_{ni} = \sum_{i=1}^n V_{ij} g_{ni} + \sum_{i=1}^n \gamma_{nij} g_{ni} - \sum_{i=1}^n \sum_{q=1}^n \omega_{nq}(W_i) V_{qj} g_{ni}$$

In Lemma A.1 we take  $r = 2$ ,  $V_k = V_{kl}$ ,  $a_{ji} = g_{nj}$ ,  $1/4 < p_1 < 1/3$  and  $p_2 = 1 - p_1$ , then

$$\left| \sum_{i=1}^n V_{ij} g_{ni} \right| = O(n^{-(2p_1-1)/2})$$

By Lemma A.2

$$\left| \sum_{i=1}^n \gamma_{nij} g_{ni} \right| \leq n \max_{i \leq n} |g_{ni}| \max_{i \leq n} |\gamma_{nij}| = o(1)$$

Use Abel's inequality and (14)

$$\left| \sum_{i=1}^n \sum_{q=1}^n \omega_{nq}(W_i) V_{qj} g_{ni} \right| \leq n \max_{i \leq n} |g_{ni}| \max_{i \leq n} \left| \sum_{q=1}^n \omega_{nq}(W_i) V_{qj} \right| = o(1).$$

The above arguments entail that  $1/\sqrt{n} \sum_{i=1}^n \tilde{X}_i g_{ni}$  is  $o(1)$ .

Observe that

$$\begin{aligned} \sum_{i=1}^n \left\{ \sum_{k=1}^n \tilde{x}_{kj} \omega_{ni}(W_k) \right\} \varepsilon_i &= \sum_{i=1}^n \left\{ \sum_{k=1}^n V_{kj} \omega_{ni}(W_k) \right\} \varepsilon_i \\ &+ \sum_{i=1}^n \left\{ \sum_{k=1}^n \gamma_{nkj} \omega_{ni}(W_k) \right\} \varepsilon_i - \sum_{i=1}^n \left[ \sum_{k=1}^n \left\{ \sum_{q=1}^n V_{qj} \omega_{nq}(W_k) \right\} \omega_{ni}(W_k) \right] \varepsilon_i \end{aligned}$$

We shall prove that all of the above three terms are  $o(n^{1/2})$ .

In Lemma A.1 we take  $r = 2$ ,  $V_k = \varepsilon_k$ ,  $a_{li} = \sum_{k=1}^n V_{kj} \omega_{ni}(W_k)$ ,  $1/4 < p_1 < 1/3$  and  $p_2 = 1 - p_1$ .

$$\left| \sum_{i=1}^n \left\{ \sum_{k=1}^n V_{kj} \omega_{ni}(W_k) \right\} \varepsilon_i \right| = O(n^{-(2p_1-1)/2} \log n)$$

By Lemma A.2 and (13), we get

$$\left| \sum_{i=1}^n \left\{ \sum_{k=1}^n \gamma_{nkj} \omega_{ni}(W_k) \right\} \varepsilon_i \right| \leq n \max_{k \leq n} \left| \sum_{i=1}^n \omega_{ni}(W_k) \varepsilon_i \right| \max_{k \leq n} |\gamma_{nkj}| = O(n^{2/3} c_n \log n)$$

Use Abel's inequality and (13) and (14), we obtain

$$\begin{aligned} \left| \sum_{i=1}^n \left[ \sum_{k=1}^n \left\{ \sum_{q=1}^n V_{qj} \omega_{nq}(W_k) \right\} \omega_{ni}(W_k) \right] \varepsilon_i \right| &\leq n \max_{k \leq n} \left| \sum_{i=1}^n \omega_{ni}(W_k) \varepsilon_i \right| \max_{k \leq n} \left| \sum_{q=1}^k V_{qj} \omega_{nq}(W_j) \right| \\ &= O(n^{1/3} \log^2 n) = o(n^{1/2}) \end{aligned}$$

We therefore complete the proof of the theorem.

## REFERENCES

- Anderson, T.W. (1984). Estimating linear statistical relationships. *Annals of Statistics* , 12, 1-45.
- Carroll, R.J., Spiegelman, C.H., Lan, K. G., Bailey, K.T. and Abbott, R.D. (1984). On errors-in-variables for binary regression models, *Biometrika* , 71, 19-25.
- Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995), *Nonlinear Measurement Error Models*. Chapman and Hall, New York.
- Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Annals of Statistics* , 16, 136-146.
- Cuzick, J. (1992a). Semiparametric additive regression. *Journal of the Royal Statistical Society, Series B* , 54, 831-843.
- Cuzick, J. (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Annals of Statistics* , 20, 1129-1136.
- Engle, R. F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* , 81, 310-320.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics* , 19, 1257-1272.
- Fan, J. & Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Annals of Statistics* , 21, 1900-1925.
- Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B* , 48, 244-248.
- Liang, H., Härdle, W. and Carroll, R.J. (1997). Large sample theory in a semiparametric partially linear errors-in-variables model. Discussion paper no. 27, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- Liang, H. and Härdle, W. (1997). Asymptotic properties of parametric estimation in partially linear heteroscedastic models. Discussion paper no. 33, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin.
- Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* , 50, 413-436.
- Stefanski, L.A. (1985). The effects of measurement error on parameter estimation. *Biometrika* , 72, 583-592.
- Stefanski, L.A. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 169-184.
- Severini, T.A. and Staniswalis, J.G. (1994). Quasilikelihood estimation in semiparametric models. *Journal of the American Statistical Association* , 89, 501-511.