

ESTIMATION IN A SEMIPARAMETRIC PARTIALLY LINEAR ERRORS-IN-VARIABLES MODEL

BY HUA LIANG,¹ WOLFGANG HÄRDLE² AND RAYMOND J. CARROLL³

*Chinese Academy of Sciences, Humboldt-Universität zu Berlin and
Texas A & M University*

We consider the partially linear model relating a response Y to predictors (X, T) with mean function $X^T\beta + g(T)$ when the X 's are measured with additive error. The semiparametric likelihood estimate of Severini and Staniswalis leads to biased estimates of both the parameter β and the function $g(\cdot)$ when measurement error is ignored. We derive a simple modification of their estimator which is a semiparametric version of the usual parametric correction for attenuation. The resulting estimator of β is shown to be consistent and its asymptotic distribution theory is derived. Consistent standard error estimates using sandwich-type ideas are also developed.

1. Introduction and background. Consider the semiparametric partially linear model based on a sample of size n ,

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

where X_i is a possibly vector-valued covariate, T_i is a scalar covariate, the function $g(\cdot)$ is unknown and the model errors ε_i are independent with conditional mean zero given the covariates. The partially linear model was introduced by Engle, Granger, Rice and Weiss (1986) to study the effect of weather on electricity demand and further studied by Heckman (1986), Chen (1988), Speckman (1988), Cuzick (1992a, b), Liang and Härdle (1997) and Severini and Staniswalis (1994).

We are interested in the estimation of the unknown parameter β and the unknown function $g(\cdot)$ in model (1) when the covariates X_i are measured

Received July 1997; revised March 1999.

¹Supported by an award from the Alexander von Humboldt Foundation and by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse."

²Supported by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse."

³Supported by National Cancer Institute Grant CA-57030, by the Texas A & M Center for Environmental and Rural Health via National Institute of Environmental Health Sciences Grant P30-ESO9106 and by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse."

AMS 1991 *subject classifications*. Primary 62J99, 62H12, 62E25, 62F10; secondary 62H25, 62F10, 62F12, 60F05.

Key words and phrases. Errors-in-variables, measurement error, nonparametric likelihood, orthogonal regression, partially linear model, semiparametric models, structural relations.

with error. Instead of observing X_i , we observe

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

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 Σ_{uu} . We will assume that Σ_{uu} is known, taking up the case that it is estimated in Section 5. The measurement error literature has been surveyed by Fuller (1987) and Carroll, Ruppert and Stefanski (1995).

If the X 's are observable, estimation of β at ordinary rates of convergence can be obtained by a local-likelihood algorithm, as follows. For every fixed β , let $\hat{g}(T, \beta)$ be an estimator of $g(T)$. For example, in the Severini and Staniswalis implementation, $\hat{g}(T, \beta)$ maximizes a weighted likelihood assuming that the model errors ε_i are homoscedastic and normally distributed, with the weights being kernel weights with symmetric kernel density function $K(\cdot)$ and bandwidth h . Having obtained $\hat{g}(T, \beta)$, β is estimated by a least squares operation,

$$\text{minimize } \sum_{i=1}^n \{Y_i - X_i^\top \beta - \hat{g}(T_i, \beta)\}^2.$$

In this particular case, the estimate for β can be determined explicitly. 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

$$(3) \quad \hat{\beta}_n = \left[\sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{X_i - \hat{g}_{x,h}(T_i)\}^\top \right]^{-1} \\ \times \sum_{i=1}^n \{X_i - \hat{g}_{x,h}(T_i)\} \{Y_i - \hat{g}_{y,h}(T_i)\}.$$

One of the important features of the estimator (3) is that it does not require undersmoothing, so that bandwidths of the usual order $h \sim n^{-1/5}$ lead to the result

$$(4) \quad n^{1/2}(\hat{\beta}_n - \beta) \Rightarrow \text{Normal}(0, B^{-1}CB^{-1}),$$

where B is the covariance matrix of $X - E(X|T)$ and C is the covariance matrix of $\varepsilon\{X - E(X|T)\}$.

The least squares form of (3) can be used to show that if one ignores the measurement error and replaces X by W , the resulting estimate is inconsistent for β . The form, though, suggests even more. It is well known that in linear regression, inconsistency caused by the measurement error can be overcome by applying the so-called "correction for attenuation." In the context of semiparametric models, this suggests that we use the estimator

$$(5) \quad \hat{\beta}_n = \left[\sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\} \{W_i - \hat{g}_{w,h}(T_i)\}^\top - n\Sigma_{uu} \right]^{-1} \\ \times \sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\} \{Y_i - \hat{g}_{y,h}(T_i)\}.$$

The estimator (5) can be derived in much the same way as the Severini–Staniswalis estimator. For every β , let $\hat{g}(T, \beta)$ maximize the weighted likelihood, ignoring the measurement error. Then form the estimators of β via a negatively penalized operation

$$(6) \quad \text{minimize } \sum_{i=1}^n \{Y_i - W_i^\top \beta - \hat{g}(T_i, \beta)\}^2 - \beta^\top \Sigma_{uu} \beta.$$

The negative sign in the second term in (6) looks odd until one remembers that the effect of the measurement error is attenuation, that is, to underestimate β in absolute value when it is scalar, and thus one must correct for attenuation by making β larger, not by shrinking it further towards zero.

In this paper, we analyze the estimate (5), and show that it is consistent, asymptotically normally distributed with a variance different from (4). Just as in the Severini–Staniswalis algorithm, the kernel weight with ordinary bandwidths of order $h \sim n^{-1/5}$ may be used.

The outline of the paper is as follows. In Section 2, we define the weighting scheme to be used and hence the estimators of β and $g(\cdot)$. Section 3 is the statement of the main results for β , while the results for $g(\cdot)$ are stated in Section 4. Section 5 states the corresponding results when the measurement error variance Σ_{uu} is estimated. Section 6 gives a numerical illustration. Final remarks are given in Section 7. All proofs are delayed until the Appendix.

2. Definition of the estimators. For technical convenience we will assume that the T_i are confined to the interval $[0, 1]$. Throughout, we shall employ $C(0 < C < \infty)$ to denote some constant not depending on n , but which may assume different values at each appearance. In our proofs and statement of results, we will let the X 's be independent random variables.

Let $\omega_{ni}(t) = \omega_{ni}(t; T_1, \dots, T_n)$ be weight functions depending only on the design points T_1, \dots, T_n . For example,

$$(7) \quad \omega_{ni}(t) = \frac{1}{h_n} \int_{s_{i-1}}^{s_i} K\left(\frac{t-s}{h_n}\right) ds, \quad 1 \leq i \leq n,$$

where $s_0 = 0$, $s_n = 1$ and $s_i = (1/2)(T_{(i)} + T_{(i+1)})$, $1 \leq i \leq n-1$, $T_{(i)}$ are the order statistics of T_i , h_n is a sequence of bandwidth parameters which tends to zero as $n \rightarrow \infty$ and $K(\cdot)$ is a nonnegative kernel function, which is supposed to have compact support and to satisfy

$$\text{supp}(K) = [-1, 1], \quad \sup |K(x)| \leq C < \infty,$$

$$\int K(u) du = 1 \quad \text{and} \quad K(u) = K(-u).$$

In this paper, for any sequence of variables or functions (S_1, \dots, S_n) , we always denote $\mathbf{S}^\top = (S_1, \dots, S_n)$, $\tilde{S}_i = S_i - \sum_{j=1}^n \omega_{nj}(T_i) S_j$, $\tilde{\mathbf{S}}^\top = (\tilde{S}_1, \dots, \tilde{S}_n)$. For example, $\tilde{\mathbf{W}}^\top = (\tilde{W}_1, \dots, \tilde{W}_n)$, $\tilde{W}_i = W_i - \sum_{j=1}^n \omega_{nj}(T_i) W_j$; $\tilde{g}_i = g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i) g(T_k)$, $\tilde{\mathbf{G}} = (\tilde{g}_1, \dots, \tilde{g}_n)^\top$.

The fact that $g(t) = E(Y_i - X_i^\top \beta | T = t) = E(Y_i - W_i^\top \beta | T = t)$ suggests

$$(8) \quad \hat{g}_n(t) = \sum_{j=1}^n \omega_{nj}(t) (Y_j - W_j^\top \hat{\beta}_n)$$

as the estimator of $g(t)$.

In some cases, it may be reasonable to assume that the model errors ε_i are homoscedastic with common variance σ^2 . In this event, since $E\{Y_i - X_i^\top \beta - g(T_i)\}^2 = \sigma^2$ and $E\{Y_i - W_i^\top \beta - g(T_i)\}^2 = E\{Y_i + X_i^\top \beta - g(T_i)\}^2 + \beta^\top \Sigma_{uu} \beta$, we define

$$(9) \quad \hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (\tilde{Y}_i - \tilde{W}_i^\top \hat{\beta}_n)^2 - \hat{\beta}_n^\top \Sigma_{uu} \hat{\beta}_n$$

as the estimator of σ^2 .

3. Main results. Let the components of X_i be $X_i = (X_{ij})$ be denoted by X_{ij} . Denote $h_j(T_i) = E(X_{ij} | T_i)$, $V_i = X_i - E(X_i | T_i)$, $1 \leq i \leq n$, $1 \leq j \leq p$. We make the following assumptions.

ASSUMPTION 1.1. $\sup_{0 \leq t \leq 1} E(\|X_1\|^4 | T = t) < \infty$ and $B = E(V_1 V_1^\top)$ is a positive definite matrix.

ASSUMPTION 1.2. $g(\cdot)$ and $h_j(\cdot)$ are Lipschitz continuous of order 1.

ASSUMPTION 1.3. The weight functions $\omega_{ni}(\cdot)$ satisfy:

- (i) $\max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) = O_P(1)$,
- (ii) $\max_{1 \leq i, j \leq n} \omega_{ni}(T_j) = O_P(b_n)$,
- (iii) $\max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_i) I(|T_j - T_i| > c_n) = O_P(c_n)$,

where $b_n = n^{-4/5}$, $c_n = n^{-1/5} \log n$.

ASSUMPTION 1.4. $E(\varepsilon_i) = E(U_i) = 0$ and $\sup_i E(\varepsilon_i^4 + \|U_i\|^4) < \infty$.

Our two main results concern the limit distributions of the estimates of β and σ^2 .

THEOREM 3.1. *Suppose that Assumptions 1.1–1.4 hold. Then $\hat{\beta}_n$ is an asymptotically normal estimator; that is,*

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

with $\Gamma = E[(\varepsilon - U^\top \beta)(X - E(X|T))]^{\otimes 2} + E\{(UU^\top - \Sigma_{uu})\beta\}^{\otimes 2} + E(UU^\top \varepsilon^2)$, where $A^{\otimes 2} = AA^\top$. Note that $\Gamma = E(\varepsilon - U^\top \beta)^2 B = E\{(UU^\top - \Sigma_{uu})\beta\}^{\otimes 2} + \Sigma_{uu} \sigma^2$ if ε is homoscedastic and independent of (X, T) .

THEOREM 3.2. *Suppose that the conditions of Theorem 3.1 hold, and that the ε 's are homoscedastic with variance σ^2 and independent of (X_i, T_i) . Then*

$$n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) \rightarrow_d N(0, \sigma_*^2),$$

where $\sigma_*^2 = E\{(\varepsilon - U^T\beta)^2 - (\beta^T \Sigma_{uu} \beta + \sigma^2)\}^2$.

REMARKS. (i) It is relatively easy to estimate the covariance matrix of $\hat{\beta}_n$. Let $\dim(X)$ be the number of the components of X . A consistent estimate of B is just

$$\{n - \dim(X)\}^{-1} \sum_{i=1}^n \{W_i - \hat{g}_{w,h}(T_i)\}^{\otimes 2} - \Sigma_{uu} =_{\text{def}} B_n.$$

In the general case, one can use (25) below to construct a consistent sandwich-type estimate of Γ , namely,

$$n^{-1} \sum_{i=1}^n \left\{ \tilde{W}_i (\tilde{Y}_i - \tilde{W}_i^T \hat{\beta}_n) + \Sigma_{uu} \hat{\beta}_n \right\}^{\otimes 2}.$$

In the homoscedastic case, namely that ε_i is independent of (X_i, T_i, U_i) with variance σ^2 and with U being normally distributed, a different formula can be used. Let $\mathcal{E}(\beta) = E\{(UU^T - \Sigma_{uu})\beta\}^{\otimes 2}$. Then a consistent estimate of Γ is

$$(\hat{\sigma}_n^2 + \hat{\beta}_n^T \Sigma_{uu} \hat{\beta}_n) \hat{B}_n + \hat{\sigma}_n^2 \Sigma_{uu} + \mathcal{E}(\hat{\beta}_n).$$

(ii) In the classical functional model [Kendall and Stuart (1992)], instead of obtaining an estimate of Σ_{uu} through replication, it is instead assumed that the ratio of Σ_{uu} to σ^2 is known. Without loss of generality, we set this ratio equal to the identity matrix. The resulting analogue of the parametric estimators to the partially linear model is to solve the following minimization problem:

$$\sum_{i=1}^n \left| \frac{\tilde{Y}_i - \tilde{W}_i^T \beta}{\sqrt{1 + \|\beta\|^2}} \right|^2 = \min!,$$

here and in the sequel $\|\cdot\|$ denotes the Euclidean norm. One can use the techniques of this paper to show that this estimator is consistent and asymptotically normally distributed. The asymptotic variance of the estimate of β for the case where ε_i is independent of (X_i, T_i) can be shown to be

$$B^{-1} \left[(1 + \|\beta\|^2)^2 \sigma^2 B + \frac{E\{(\varepsilon - U^T\beta)^2 \Gamma_1 \Gamma_1^T\}}{1 + \|\beta\|^2} \right] B^{-1},$$

where $\Gamma_1 = (1 + \|\beta\|^2)U + (\varepsilon - U^T\beta)\beta$.

4. Asymptotic results for the nonparametric part.

THEOREM 4.1. *Suppose that Assumptions 1.1–1.4 hold and that $\omega_{n_i}(t)$ are Lipschitz continuous of order 1 for all $i = 1, \dots, n$. Then for fixed T_i , the asymptotic bias and asymptotic variance of $\hat{g}_n(t)$ are, respectively, $\sum_{i=1}^n \omega_{n_i}(t)g(T_i) - g(t)$ and $\sum_{i=1}^n \omega_{n_i}^2(t)(\beta^\top \Sigma_{uu} \beta + \sigma^2)$. These are all of order $O(n^{-2/5})$ for the kernel estimators.*

5. Estimated error variance. Although in some cases the measurement error covariance matrix Σ_{uu} has been established by independent experiments, in others it is unknown and must be estimated. The usual method of doing so [Carroll, Ruppert and Stefanski (1995), Chapter 3] is by partial replication, so that we observe $W_{ij} = X_i + U_{ij}$, $j = 1, \dots, m_i$.

For notational convenience, we consider here only the case that $m_i \leq 2$ and assume that a fraction δ of the data has such replicates. Let \bar{W}_i be the sample mean of the replicates. Then a consistent, unbiased method of moments estimate for Σ_{uu} is

$$\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} (W_{ij} - \bar{W}_i)(W_{ij} - \bar{W}_i)^\top}{\sum_{i=1}^n (m_i - 1)}.$$

The estimator changes only slightly to accommodate the replicates, becoming

$$\begin{aligned} \hat{\beta}_n &= \left[\sum_{i=1}^n \{ \bar{W}_i - \hat{g}_{w,h}(T_i) \}^{\otimes 2} - n(1 - \delta/2) \hat{\Sigma}_{uu} \right]^{-1} \\ (10) \quad &\times \sum_{i=1}^n \{ \bar{W}_i - \hat{g}_{w,h}(T_i) \} \{ Y_i - \hat{g}_{y,h}(T_i) \}, \end{aligned}$$

where $\hat{g}_{w,h}(\cdot)$ is the kernel regression of the \bar{W}_i 's on T_i .

Using the techniques in the Appendix, one can show that the limit distribution of (10) is Normal(0, $B^{-1} \Gamma_2 B^{-1}$), with

$$\begin{aligned} \Gamma_2 &= (1 - \delta) E \left[(\varepsilon - U^\top \beta) \{ X - E(X|T) \} \right]^{\otimes 2} \\ &+ \delta E \left[(\varepsilon - \bar{U}^\top \beta) \{ X - E(X|T) \} \right]^{\otimes 2} \\ (11) \quad &+ (1 - \delta) E \left[\{ UU^\top - (1 - \delta/2) \Sigma_{uu} \} \beta \right]^{\otimes 2} + UU^\top \varepsilon^2 \\ &+ \delta E \left[\left\{ \{ \bar{U}\bar{U}^\top - (1 - \delta/2) \Sigma_{uu} \} \beta \right\}^{\otimes 2} + \bar{U}\bar{U}^\top \varepsilon^2 \right]. \end{aligned}$$

In (11), \bar{U} refers to the mean of two U 's. In the case that ε is independent of (X, T) , the sum of the first two terms simplifies to $\{ \sigma^2 + \beta^\top (1 - \delta/2) \Sigma_{uu} \beta \} B$.

Standard error estimates can also be derived. A consistent estimate of B is

$$\hat{B}_n = \{ n - \dim(X) \}^{-1} \sum_{i=1}^n \{ \bar{W}_i - \hat{g}_{w,h}(T_i) \}^{\otimes 2} - (1 - \delta/2) \hat{\Sigma}_{uu}.$$

Estimates of Γ_2 can also be easily developed. In the homoscedastic case with normal errors, the sum of the first two terms can be estimated by $(\hat{\sigma}_n^2 + (1 - \delta/2)\hat{\beta}_n^\top \hat{\Sigma}_{uu} \hat{\beta}_n) \hat{B}_n$. The sum of the last two terms is a deterministic function of $(\beta, \sigma^2, \Sigma_{uu})$, and these estimates are simply substituted into the formula.

A general sandwich-type estimator is developed as follows. Define $\kappa = n^{-1} \sum_{i=1}^n m_i^{-1}$, and define

$$R_i = \tilde{W}_i (\tilde{Y}_i - \tilde{W}_i^\top \hat{\beta}_n) + \frac{\hat{\Sigma}_{uu} \hat{\beta}_n}{m_i} + \frac{\kappa}{\delta} (m_i - 1) \left\{ \frac{1}{2} (W_{i1} - W_{i2})(W_{i1} - W_{i2})^\top - \hat{\Sigma}_{uu} \right\}.$$

Then a consistent estimate of Γ_2 is the sample covariance matrix of the R_i 's.

6. Numerical example. To illustrate our method, we consider data from the Framingham Heart Study. We consider $n = 1615$ males with Y being their average blood pressure in a fixed two-year period, T being their age and W being the logarithm of the observed cholesterol level, for which there are two replicates.

We do two analyses. In the first, we use both cholesterol measurements, so that in the notation of Section 5, $\delta = 1$. In this analysis, there is not a great deal of measurement error. Thus, in our second analysis, which is given for illustrative purposes, we use only the first cholesterol measurement, but fix the measurement error variance at the value obtained in the first analysis, in which case $\delta = 0$. For nonparametric fitting, we chose the bandwidth using cross-validation to predict the response. In precise terms, we compute the squared error using a geometric sequence of 191 bandwidths ranging in $[1, 20]$. The optimal bandwidth is selected to minimize the squared error among these 191 candidates. An analysis ignoring the measurement error found some curvature in T ; see Figure 1 for the estimate of $g(T)$. All calculations were performed in XploRe [Härdle, Klinkle and Turlach (1995)].

Our results are as follows. First, consider the case that the measurement error is estimated and both cholesterol values are used to estimate Σ_{uu} . The estimator of β ignoring the measurement error is 9.438, with estimated standard error 0.187. When we account for the measurement error, the estimate increases to $\hat{\beta} = 12.540$ and the standard error increases to 0.195.

In the second analysis, we fix the measurement error variance and use only the first cholesterol value. The estimator of β ignoring the measurement error was 10.744, with estimated standard error 0.492. When we account for the measurement error, the estimate increases to $\hat{\beta} = 13.690$ and the standard error increases to 0.495.

7. Discussion. The nonparametric regression estimator (8) is based on locally weighted averages. Clearly, results such as Theorem 3.1 should apply if (8) is replaced by a locally linear kernel regression estimator or by a spline estimator, although our proofs do not apply to these estimators.

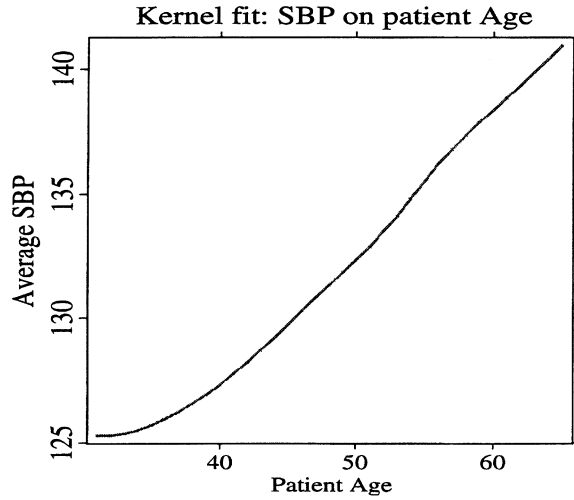


FIG. 1. Estimate of the function $g(T)$ in the Framingham data ignoring measurement error.

We have treated the case that the parametric part X of the model has measurement error and the nonparametric part T is measured exactly. An interesting problem is to interchange the roles of X and T , so that the parametric part is measured exactly and the nonparametric part is measured with error, that is, $E(Y|X, T) = \theta T + g(X)$. Fan and Truong (1993) have shown in this case that with normally distributed measurement error, the nonparametric function $g(\cdot)$ can be estimated only at logarithmic rates and not with rate $n^{-2/5}$. We conjecture even so that θ can be estimated at parametric rates, but this remains an open problem.

APPENDIX

In this Appendix, we prove several required lemmas. Lemma A.1 provides bounds for $h_j(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)h_j(T_k)$ and $g(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)g(T_k)$. The proof is immediate.

LEMMA A.1. Suppose that Assumptions 1.1–1.4 hold. Then

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

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

LEMMA A.2. If Assumptions 1.1–1.4 hold, then $n^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = B + o_p(1)$.

PROOF. Denote $\bar{h}_{ns}(T_i) = h_s(T_i) - \sum_{k=1}^n \omega_{nk}(T_i)X_{ks}$. It follows from $X_{js} = h_s(T_j) + V_{js}$ that the (s, m) th element of $\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}}$ ($s, m = 1, \dots, p$) is

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

The strong law of large numbers implies that $n^{-1} \sum_{i=1}^n V_i V_i^\top = B + o_p(1)$, and Lemma A.1 means $R_{nsm}^{(3)} = o_p(n)$, which together with the Cauchy-Schwarz inequality shows that $R_{nsm}^{(1)} = o_p(n)$ and $R_{nsm}^{(2)} = o_p(n)$. This completes the proof of the lemma. \square

LEMMA A.3 (Bernstein's inequality). *Let $\Gamma_1, \dots, \Gamma_n$ be independent random variables with zero means and bounded ranges, $|\Gamma_i| \leq M$. Then for each $\eta > 0$,*

$$P\left\{\left|\sum_{i=1}^n \Gamma_i\right| > \eta\right\} \leq 2 \exp\left\{-\eta^2 / \left[2\left\{\sum_{i=1}^n \text{var}(\Gamma_i) + M\eta\right\}\right]\right\}.$$

Denote $\varepsilon'_j = \varepsilon_j I(|\varepsilon_j| \leq n^{1/4})$ and $\varepsilon''_j = \varepsilon_j - \varepsilon'_j = \varepsilon_j I(|\varepsilon_j| > n^{1/4})$, $j = 1, \dots, n$. We next establish several results for nonparametric regression.

LEMMA A.4. *Assume that Assumptions 1.3 and 1.4 hold. Then*

$$\max_{1 \leq i \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_i) \varepsilon_k \right| = o_p\{n^{-2/5} \log(n)\}.$$

PROOF. Fix $L > 0$ but arbitrarily large. Let

$$B_{nL} = \left\{ \max_{1 \leq i \leq n} \sum_{j=1}^n w_{nj}(T_i) \leq L, \max_{1 \leq i, j \leq n} w_{nj}(T_i) \leq Lb_n \right\}.$$

Then

$$\begin{aligned} &P\left\{\max_{1 \leq i \leq n} \left| \sum_{j=1}^n w_{nj}(T_i) \varepsilon_j \right| > n^{-2/5} \log(n)\right\} \\ (12) \quad &\leq P\{I(B_{nL}) = 0\} \\ &\quad + P\left\{\max_{1 \leq i \leq n} \left| \sum_{j=1}^n w_{nj}(T_i) \varepsilon_j \right| > n^{-2/5} \log(n), I(B_{nL}) = 1\right\}. \end{aligned}$$

Since by Assumption 1.3, $P\{I(B_{nL}) = 1\}$ can be made arbitrarily small by choosing L sufficiently large, it suffices to show that the second term in (12) converges to zero for any L .

Application of Bernstein's inequality to (12) is complicated by the fact that the terms $w_{n_j}(T_i)$ and $I(B_{nL}) = 1$ are random. We first condition on these terms and will later uncondition. For sufficiently large C , first note that

$$\begin{aligned} &P\left\{\max_{1 \leq i \leq n} \left| \sum_{j=1}^n w_{n_j}(T_i) \{\varepsilon'_j - E(\varepsilon'_j)\} \right| \right. \\ &\quad \left. > Cn^{-2/5} \log(n) | \{w_{n_j}(T_i)\}, I(B_{nL}) = 1 \right\} \\ &\leq \sum_{i=1}^n P\left\{ \left| \sum_{j=1}^n w_{n_j}(T_i) \{\varepsilon'_j - E(\varepsilon'_j)\} \right| \right. \\ &\quad \left. > Cn^{-2/5} \log(n) | \{w_{n_j}(T_i)\}, I(B_{nL}) = 1 \right\}. \end{aligned}$$

Now apply Bernstein's inequality with $\eta = Cn^{-2/5} \log(n)$ and $M = 2Lb_n n^{1/4}$. Then the right-hand side of the last expression is bounded by

$$(13) \quad 2I(B_{nL}) \sum_{i=1}^n \exp\left\{ - \frac{C^2 n^{-4/5} \log^2(n)}{4LCb_n n^{1/4-2/5} \log(n) + 2\sum_{j=1}^n w_{n_j}^2(T_i) \text{var}(\varepsilon'_j)} \right\}.$$

First note that $b_n = n^{-4/5}$ and $\text{var}(\varepsilon'_j) < \infty$. On the set that $I(B_{nL}) = 1$, we have thus that

$$\sum_{j=1}^n w_{n_j}^2(T_i) \leq \sum_{j=1}^n w_{n_j}(T_i) \max_{1 \leq i, j \leq n} w_{n_j}(T_i) \leq L^2 b_n.$$

This means that (13) is bounded by $2nI(B_{nL}) \exp\{- (C/L) \log(n)\} \leq n^{-3/2}$ for sufficiently large C . Since this last expression is independent of the $\{w_{n_j}(T_i)\}$ except through $I(B_{nL})$, we have that

$$P\left\{ \max_{1 \leq i \leq n} \left| \sum_{j=1}^n w_{n_j}(T_i) \{\varepsilon'_j - E(\varepsilon'_j)\} \right| > Cn^{-2/5} \log(n) | I(B_{nL}) = 1 \right\} \leq n^{-3/2}.$$

This shows that

$$(14) \quad \max_{1 \leq i \leq n} \left| \sum_{j=1}^n w_{n_j}(T_i) \{\varepsilon'_j - E(\varepsilon'_j)\} \right| = o_p\{n^{-2/5} \log(n)\}.$$

Now consider $V_n = \max_{1 \leq i \leq n} \sum_{j=1}^n w_{n_j}(T_i) \{\varepsilon''_j - E(\varepsilon''_j)\}$. Let p and q be such that $1 \leq p < 2$, $1/p + 1/q = 1$ and $1/q < 2/5 - 1/4$. By Hölder's inequality,

$$|V_n| \leq \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^n w_{n_j}^q(T_i) \right\}^{1/q} \left\{ \sum_{j=1}^n |\varepsilon''_j - E(\varepsilon''_j)|^p \right\}^{1/p}.$$

By Assumption 1.3(ii), $w_{nj}^q(T_i) = O_p(b_n^q)$ so that $\sum_j w_{nj}^q(T_i) = O_p(nb_n^q) = O_p(n^{1-4q/5})$, and thus

$$|V_n| \leq O_p\{n^{(1-4q/5)/q}\} \left\{ \sum_{j=1}^n |\varepsilon_j'' - E(\varepsilon_j'')|^p \right\}^{1/p}.$$

Clearly,

$$(15) \quad n^{-1} \sum_{j=1}^n \left[|\varepsilon_j'' - E(\varepsilon_j'')|^p - E\left\{ |\varepsilon_j'' - E(\varepsilon_j'')|^p \right\} \right] = o_p(1).$$

Also, again using Hölder's inequality,

$$E|\varepsilon_j''|^p = E\left\{ |\varepsilon_j|^{pI(\varepsilon_j > n^{1/4})} \right\} \leq \left(E|\varepsilon_j|^4 \right)^{p/4} \left\{ P(|\varepsilon_j| > n^{1/4}) \right\}^{1-p/4},$$

which by Chebyshev's inequality is bounded by $\leq n^{-1+p/4} (E|\varepsilon_j|^4)^{p/4}$. It thus follows that

$$(16) \quad \sum_{j=1}^n E|\varepsilon_j'' - E(\varepsilon_j'')|^p = O_p(n^{p/4}).$$

Replacing (16) into (15), we get

$$\sum_{j=1}^n |\varepsilon_j'' - E(\varepsilon_j'')|^p = O_p(n^{p/4}),$$

where, along with the fact that $1/q < 2/5 - 1/4$, we find that

$$\max_{1 \leq i \leq n} \sum_{j=1}^n w_{nj}(T_i) \{ \varepsilon_j'' - E(\varepsilon_j'') \} = O_p(n^{(1-4q/5)/q+1/4}) = o_p(n^{-2/5}).$$

This completes the proof of Lemma A.4. \square

LEMMA A.5. *Suppose that Assumptions 1.1–1.4 hold. Then*

$$\begin{aligned} \sum_{i=1}^n U_i \tilde{g}_i &= o_p(n^{1/2}), \\ \sum_{i=1}^n \varepsilon_i \tilde{g}_i &= o_p(n^{1/2}). \end{aligned}$$

The same holds if $g(T_i)$ is replaced by $h_j(T_i)$.

PROOF. We prove only the first step, as the other steps follow in a similar fashion. Let $\xi_n = n^{1/2}/\log(n)$:

$$\begin{aligned} P\left(\left| \sum_{i=1}^n U_i \tilde{g}_i \right| > \xi_n \right) &\leq P\left(\left| \sum_{i=1}^n U_i \tilde{g}_i \right| > \xi_n, \max_i |\tilde{g}_i| \leq c_n \log n \right) \\ &\quad + P\left(\max_i |\tilde{g}_i| > c_n \log n \right). \end{aligned}$$

The second term is $o_p(1)$ by Lemma A.1. For the first term, let r_i be the event that $|\tilde{g}_i| \leq c_n \log(n)$. Then,

$$\begin{aligned}
 & P \left[\left| \sum_{i=1}^n U_i \tilde{g}_i \right| > \xi_n, \{I(r_i) = 1 \forall i\} \right] \\
 (17) \quad & \leq \xi_n^{-2} \sum_{i=1}^n E[U_i \tilde{g}_i \{I(r_i) = 1\}]^2 \\
 & \quad + \xi_n^{-2} \sum_{i \neq k}^n E[U_i U_k \tilde{g}_i \tilde{g}_k I(r_k) = 1 \forall k].
 \end{aligned}$$

Since $\tilde{g}_i \{I(r_i) = 1\} \leq c_n \log(n)$ is independent of U_i , the first term in (17) is $O\{n \xi_n^{-2} c_n^2 \log^2(n)\} = o(1)$. The second term is easily seen to equal zero. \square

LEMMA A.6. *Suppose that Assumptions 1.1–1.4 hold. Then*

$$\begin{aligned}
 n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j U_i &= o_p(1), \\
 n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j \varepsilon_i &= o_p(1), \\
 n^{-1/2} \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) U_j U_i &= o_p(1).
 \end{aligned}$$

PROOF. We prove only the first step, as the other steps follow in a similar fashion. Let r_{ij} be the event that $|w_{nj}(T_i)| \leq Cb_n \log n$:

$$\begin{aligned}
 & P \left\{ n^{-1/2} \left| \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j U_i \right| > \xi \right\} \\
 & \leq P \left\{ n^{-1/2} \left| \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j U_i \right| > \xi, I(r_{ij} = 1 \forall i, j) \right\} \\
 & \quad + P \left\{ \max_{i,j} |w_{nj}(T_i)| > Cb_n \log n \right\}.
 \end{aligned}$$

The second term tends to zero by Assumption 1.3(ii). For the first term, note that

$$\begin{aligned}
 & P \left\{ n^{-1/2} \left| \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j U_i \right| > \xi, I(r_{ij} = 1 \forall i, j) \right\} \\
 & \leq n^{-1} \xi^{-2} E \left\{ \sum_{i=1}^n \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j U_i I(r_{ij} = 1 \forall i, j) \right\}^2 \\
 & = n^{-1} \xi^{-2} \sum_{i=1}^n E \left\{ \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j I(r_{ij} = 1 \forall i, j) \right\}^2 E U_i^2.
 \end{aligned}$$

The last equation holds because U_i and $\sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j I(r_{ij} = 1 \forall i, j)$ are independent for each i , and U_i are iid with mean zero. It suffices to prove

$$\max_i E \left\{ \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j I(r_{ij} = 1 \forall i, j) \right\}^2 \rightarrow 0.$$

In fact,

$$\begin{aligned} & E \left\{ \sum_{j=1}^n \omega_{nj}(T_i) \varepsilon_j I(r_{ij} = 1 \forall i, j) \right\}^2 \\ &= \sum_{j=1}^n E \left\{ \omega_{nj}(T_i) \varepsilon_j I(r_{ij} = 1 \forall i, j) \right\}^2 \\ &\quad + \sum_{j \neq k}^n E \left\{ \omega_{nj}(T_i) \varepsilon_j \omega_{nk}(T_i) \varepsilon_k I(r_{ij} = 1 \forall i, j) \right\}. \end{aligned}$$

The second term equals zero. The first term equals

$$\sum_{j=1}^n E \left[\left\{ \omega_{nj}(T_i) \varepsilon_j \right\}^2 \{ I(r_{ij} = 1 \forall i, j) \} \right],$$

and this is $O(nb_n^2 \log^2(n)) = o(1)$, as required. \square

LEMMA A.7. *Assume that Assumptions 1.1–1.4 hold. Then*

$$(18) \quad p \lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{W}}^\top \tilde{\mathbf{W}} = B + \Sigma_{uu},$$

$$(19) \quad p \lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{W}}^\top \tilde{\mathbf{Y}} = B\beta,$$

$$(20) \quad p \lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{Y}}^\top \tilde{\mathbf{Y}} = \beta^\top B\beta + \sigma^2.$$

PROOF. Since $W_i = X_i + U_i$ and $\tilde{W}_i = \tilde{X}_i + \tilde{U}_i$, for the (s, m) matrix element we obtain

$$(21) \quad \begin{aligned} n^{-1} (\tilde{\mathbf{W}}^\top \tilde{\mathbf{W}})_{sm} &= n^{-1} (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}})_{sm} + n^{-1} (\tilde{\mathbf{U}}^\top \tilde{\mathbf{X}})_{sm} \\ &\quad + n^{-1} (\tilde{\mathbf{X}}^\top \tilde{\mathbf{U}})_{sm} + n^{-1} (\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}})_{sm}. \end{aligned}$$

First, we prove that the second and third terms converge to zero. It follows from the strong law of large numbers and Lemma A.2 that

$$(22) \quad n^{-1} \sum_{j=1}^n X_{js} U_{jm} \rightarrow 0 \quad \text{a.s.}$$

Observe that

$$\begin{aligned} n^{-1}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{U}})_{sm} &= n^{-1} \left[\sum_{j=1}^n X_{js} U_{jm} - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} U_{jm} \right. \\ &\quad \left. - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} X_{js} \right. \\ &\quad \left. + \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} \right]. \end{aligned}$$

Similarly to the proof of Lemma A.4, we can prove that

$$\sup_{1 \leq j \leq n} \left| \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right| = o_P(1),$$

which, together with (22) and Assumption 1.3(ii), imply that each term above tends to zero. The same reason implies that $n^{-1}(\tilde{\mathbf{U}}^\top \tilde{\mathbf{X}})_{sm}$ also tends to zero.

Second, we prove

$$(23) \quad n^{-1}(\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}})_{sm} \rightarrow \sigma_{sm}^2,$$

where σ_{sm}^2 is the (s, m) th element of Σ_{uu} ,

$$\begin{aligned} n^{-1}(\tilde{\mathbf{U}}^\top \tilde{\mathbf{U}})_{sm} &= n^{-1} \left[\sum_{j=1}^n U_{js} U_{jm} - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{ks} \right\} U_{jm} \right. \\ &\quad \left. - \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} U_{js} \right. \\ &\quad \left. + \sum_{j=1}^n \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{ks} \right\} \left\{ \sum_{k=1}^n \omega_{nk}(T_j) U_{km} \right\} \right]. \end{aligned}$$

Obviously, $n^{-1} \sum_{j=1}^n U_{js} U_{jm} \rightarrow \sigma_{sm}^2$. It follows from Lemmas A.4 and A.6 that (23) holds. Using (21), (23) and the arguments for $n^{-1}(\tilde{\mathbf{U}}^\top \tilde{\mathbf{X}})_{sm} \rightarrow 0$ and $n^{-1}(\tilde{\mathbf{X}}^\top \tilde{\mathbf{U}})_{sm} \rightarrow 0$, we complete the proof of (18).

We now prove (19). Note that $\tilde{\mathbf{W}}^\top \tilde{\mathbf{Y}} = \tilde{\mathbf{W}}^\top (\tilde{\mathbf{X}}\beta + \tilde{\mathbf{G}} + \tilde{\varepsilon})$. From Lemma 1, $\sum_{j=1}^n \tilde{g}_j^2 = O_P(c_n^2 n)$, so that

$$\left| \sum_{j=1}^n X_{js} \tilde{g}_j \right| \leq \left(\sum_{j=1}^n X_{js}^2 \sum_{j=1}^n \tilde{g}_j^2 \right)^{1/2} \leq O_P(c_n n^{1/2}) \left(\sum_{j=1}^n X_{js}^2 \right)^{1/2} = O_P(Cnc_n)$$

and

$$\begin{aligned} (\tilde{\mathbf{W}}^\top \tilde{\mathbf{G}})_s &= \sum_{j=1}^n \tilde{X}_{js} \tilde{g}_j + \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j \\ &= \sum_{j=1}^n \left\{ X_{js} - \sum_{k=1}^n \omega_{nk}(T_j) X_{ks} \right\} \tilde{g}_j + \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j. \end{aligned}$$

Obviously, $n^{-1} \sum_{j=1}^n \tilde{U}_{js} \tilde{g}_j$ tends to zero. Therefore $n^{-1}(\tilde{\mathbf{W}}^\top \tilde{\mathbf{G}})_s$ tends to zero.

The proof that $n^{-1}(\tilde{\mathbf{W}}^\top \tilde{\boldsymbol{\varepsilon}})_s$ tends to zero is similar to that of $n^{-1}(\tilde{\mathbf{W}}^\top \tilde{\mathbf{U}})_s \rightarrow 0$. Combining the above arguments and (18), we complete the proof of (19). The proof of (20) can be completed by similar arguments. The details are omitted. \square

LEMMA A.8. *Assume that Assumptions 1.1–1.4 hold. Then*

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \tilde{\boldsymbol{\varepsilon}}_i \tilde{X}_i &= n^{-1/2} \sum_{i=1}^n \boldsymbol{\varepsilon}_i V_i + o_P(1), \\ n^{-1/2} \sum_{i=1}^n \tilde{X}_i \tilde{U}_i^\top &= n^{-1/2} \sum_{i=1}^n V_i U_i^\top + o_P(1). \end{aligned}$$

PROOF. We show only the first step, as the second step follows in a similar fashion. Let $h(T) = E(X|T)$ and $h_i = h(T_i)$. By a direct calculation,

$$n^{-1/2} \sum_{i=1}^n \boldsymbol{\varepsilon}_i (V_i - \tilde{X}_i) = n^{-1/2} \sum_{i=1}^n \boldsymbol{\varepsilon}_i \tilde{h}_i - n^{-1/2} \sum_{i=1}^n \boldsymbol{\varepsilon}_i \sum_{j=1}^n w_{nj}(T_i) \{X_j - h(T_j)\}.$$

The first term is $o_P(1)$ by Lemma A.4. The second term follows, using Assumption 1.1 by using the same method of proof as in Lemma A.6, upon remembering that for $j \neq k$,

$$E[\{X_j - h(T_j)\}\{X_k - h(T_k)\}|T_1, \dots, T_n] = 0. \quad \square$$

PROOF OF THEOREM 3.1. Denote $\Delta_n = (\tilde{\mathbf{W}}^\top \tilde{\mathbf{W}} - n\Sigma_{uu})/n$. By Lemma A.7 and a direct calculation,

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) &= n^{1/2} \Delta_n^{-1} (\tilde{\mathbf{W}}^\top \tilde{\mathbf{Y}} + \tilde{\mathbf{W}}^\top \tilde{\mathbf{W}} \boldsymbol{\beta} + n\Sigma_{uu} \boldsymbol{\beta}) \\ &= n^{-1/2} \Delta_n^{-1} (\tilde{\mathbf{X}}^\top \tilde{\mathbf{G}} + \tilde{\mathbf{X}}^\top \tilde{\boldsymbol{\varepsilon}} + \tilde{\mathbf{U}}^\top \tilde{\mathbf{G}} + \tilde{\mathbf{U}}^\top \tilde{\boldsymbol{\varepsilon}} \\ &\quad - \tilde{\mathbf{X}}^\top \tilde{\mathbf{U}} \boldsymbol{\beta} - \tilde{\mathbf{U}}^\top \tilde{\mathbf{U}} \boldsymbol{\beta} + n\Sigma_{uu} \boldsymbol{\beta}). \end{aligned}$$

By Lemmas A.1–A.2, A.4–A.6 and A.8, it is an easy calculation to show that

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) &= n^{-1/2} \Delta_n^{-1} \\ (24) \quad &\times \sum_{i=1}^n (V_i \boldsymbol{\varepsilon}_i - V_i U_i^\top \boldsymbol{\beta} + U_i \boldsymbol{\varepsilon}_i - U_i U_i^\top \boldsymbol{\beta} + \Sigma_{uu} \boldsymbol{\beta}) \\ &+ o_P(1) \end{aligned}$$

$$(25) \quad =_{\text{def}} n^{-1/2} \sum_{i=1}^n \zeta_{in} + o_P(1).$$

Since $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i = \mathbf{0}$ and $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n V_i V_i^\top = \mathbf{B}$ and $\sup_i E(\boldsymbol{\varepsilon}_i^4 + \|U\|^4) < \infty$, it follows that the sequence of k th elements $\{\zeta_{in}^{(k)}\}$ of $\{\zeta_{in}\}$ ($k = 1, \dots, p$) satisfy, for any given $\zeta > 0$, $n^{-1} \sum_{i=1}^n E\{\zeta_{in}^{(k)2} I(|\zeta_{in}^{(k)}| > \zeta n^{1/2})\} \rightarrow 0$ as $n \rightarrow \infty$. This means that the Lindeberg condition for the central limit theorem

holds. Moreover, note that

$$\begin{aligned} \text{cov}(\zeta_{ni}) &= E\{V_i(\varepsilon_i - U_i^\top\beta)\}^{\otimes 2} + E\{(U_iU_i^\top - \Sigma_{uu})\beta\}^{\otimes 2} + E(U_iU_i^\top\varepsilon_i^2) \\ &\quad + E(V_iU_i^\top\beta\beta^\top U_iU_i^\top) + E(U_iU_i^\top\beta\beta^\top U_i)V_i. \end{aligned}$$

These arguments ensure that

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \text{cov}(\zeta_{ni}) &= E[(\varepsilon - U^\top\beta)\{X - E(X|T)\}]^{\otimes 2} \\ &\quad + E\{(UU^\top - \Sigma_{uu})\beta\}^{\otimes 2} + E(UU^\top\varepsilon^2). \end{aligned}$$

Theorem 3.1 now follows. \square

PROOF OF THEOREM 3.2. Denote

$$\begin{aligned} A_n &= n^{-1} \begin{bmatrix} \tilde{\mathbf{Y}}^\top \tilde{\mathbf{Y}} & \tilde{\mathbf{Y}}^\top \tilde{\mathbf{W}} \\ \tilde{\mathbf{W}}^\top \tilde{\mathbf{Y}} & \tilde{\mathbf{W}}^\top \tilde{\mathbf{W}} \end{bmatrix}; \quad A = \begin{bmatrix} \beta^\top B\beta + \sigma^2 & \beta^\top B \\ B\beta & B + \Sigma_{uu} \end{bmatrix}; \\ \tilde{A}_n &= n^{-1} \begin{bmatrix} (\varepsilon + \mathbf{V}\beta)^\top (\varepsilon - \mathbf{V}\beta) & (\varepsilon + \mathbf{V}\beta)^\top (\mathbf{U} + \mathbf{V}) \\ (\mathbf{U} + \mathbf{V})^\top (\varepsilon + \mathbf{V}\beta) & (\mathbf{U} + \mathbf{V})^\top (\mathbf{U} + \mathbf{V}) \end{bmatrix}. \end{aligned}$$

Note that $\hat{\sigma}_n^2 = (1, -\hat{\beta}_n^\top)A_n(1, -\tilde{\beta}_n^\top)^\top - \tilde{\beta}_n^\top \Sigma_{uu} \tilde{\beta}_n^\top$. A direct calculation using Lemma A.6 yields that $n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) = n^{1/2} \sum_{j=1}^4 S_{jn} + n^{-1/2}(\varepsilon - \mathbf{U}\beta)^\top (\varepsilon - \mathbf{U}\beta) - n^{1/2}(\beta^\top \Sigma_{uu} \beta + \sigma^2) + o_p(1)$, where $S_{1n} = (1, -\hat{\beta}_n^\top)(A_n - \tilde{A}_n)(1, -\hat{\beta}_n^\top)^\top$, $S_{2n} = (1, -\hat{\beta}_n^\top)(\tilde{A}_n - A)(0, \beta^\top - \hat{\beta}_n^\top)^\top$, $S_{3n} = (0, \beta^\top - \hat{\beta}_n^\top)(\tilde{A}_n - A)(1, -\beta^\top)^\top$, $S_{4n} = -(\beta - \hat{\beta}_n)^\top B(\beta - \hat{\beta}_n)$. It follows from Theorem 3.1 and Lemma A.7 that $n^{1/2}S_{jn}$ converges to zero in probability for $j = 2, 3, 4$. To show that $n^{1/2}S_{1n} = o_p(1)$ is more detailed, but follows from Lemmas A.1, A.4–A.6. This means that

$$n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) = n^{-1/2} \sum_{i=1}^n \left\{ (\varepsilon_i - U_i^\top\beta)^2 - (\beta^\top \Sigma_{uu} \beta + \sigma^2) \right\} + o_p(1).$$

Theorem 3.2 now follows immediately. \square

PROOF OF THEOREM 4.1. Since $\hat{\beta}_n$ is a consistent estimator of β , its asymptotic bias and variance equal the relative ones of $\sum_{j=1}^n \omega_{nj}(t)(Y_j - W_j^\top\beta)$, which is denoted by $\hat{g}_n^*(t)$. By a simple calculation,

$$\begin{aligned} E\hat{g}_n^*(t) - g(t) &= \sum_{i=1}^n \omega_{ni}(t)g(T_i) - g(t), \\ \hat{g}_n^*(t) - E\hat{g}_n^*(t) &= \sum_{i=1}^n \omega_{ni}^2(t)(\beta^\top \Sigma_{uu} \beta + \sigma^2). \end{aligned}$$

Both terms are $O(n^{-2/5})$ by Lemma A.1 and Assumption 1.3(iii). Theorem 4.1 then follows. \square

Acknowledgments. The authors are extremely grateful to the Editor, Associate Editor and an anonymous referee for their many valuable suggestions and comments, which greatly improved the presentation of the paper.

REFERENCES

- 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. *Ann. Statist.* **16** 136–146.
- CUZICK, J. (1992a). Semiparametric additive regression. *J. Roy. Statist. Soc. Ser. B* **54** 831–843.
- CUZICK, J. (1992b). Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.* **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. *J. Amer. Statist. Assoc.* **81** 310–320.
- FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925.
- FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- HÄRDLE, W., KLINKE, S. and TURLACH, B. A. (1995). *XploRe: An Interactive Statistical Computing Environment*. Springer, New York.
- HECKMAN, N. E. (1986). Spline smoothing in partly linear models. *J. Roy. Statist. Soc. Ser. B* **48** 244–248.
- KENDALL, M. and STUART, A. (1992). *The Advanced Theory of Statistics* **2**, 4th ed. Griffin, London.
- LIANG, H. and HÄRDLE, W. (1997). Asymptotic normality of parametric part in partially linear heteroscedastic regression models. DP 33, SFB 373, Humboldt Univ. Berlin.
- SEVERINI, T. A. and STANISWALIS, J. G. (1994). Quasilikelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89** 501–511.
- SPECKMAN, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50** 413–436.

H. LIANG
 INSTITUTE OF SYSTEMS SCIENCE
 CHINESE ACADEMY OF SCIENCES
 BEIJING 100080
 CHINA
 E-MAIL: hliang@iss01.iss.ac.cn

W. HÄRDLE
 INSTITUT FÜR STATISTIK UND ÖKONOMETRIE
 HUMBOLDT-UNIVERSITÄT ZU BERLIN
 D-10178 BERLIN
 GERMANY
 E-MAIL: haerdle@wiwi.hu-berlin.de

R. J. CARROLL
 DEPARTMENT OF STATISTICS
 TEXAS A & M UNIVERSITY
 COLLEGE STATION, TEXAS 77843-3143
 E-MAIL: carroll@stat.tamu.edu