Liang, Hua; Wang, Naisyin

**Working Paper**

# Theoretical properties of two estimators in partially linear single-index measurement error models

SFB 373 Discussion Paper, No. 1999,104

**Provided in Cooperation with:**
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

*Suggested Citation:* Liang, Hua; Wang, Naisyin (1999) : Theoretical properties of two estimators in partially linear single-index measurement error models, SFB 373 Discussion Paper, No. 1999,104, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10046906

This Version is available at:
https://hdl.handle.net/10419/61757

# THEORETICAL PROPERTIES OF TWO ESTIMATORS IN PARTIALLY LINEAR SINGLE-INDEX MEASUREMENT ERROR MODELS *

Hua Liang and Naisyin Wang
Department of Statistics
Texas A&M University
College Station
TX 77843-3143, USA

**Abstract**

Considering partially linear single-index errors-in-variables model which can be described as $Y = \eta(X^T\alpha_0) + Z^T\beta_0 + \varepsilon$ when the $Z'$s are measured with additive errors. The general estimators established in literature are biased when ignoring the measurement errors. We proposed two estimators in this setting. Their theoretical properties were derived and compared.

**AMS 1991 subject classification:** Primary: 62G07, 62G05; Secondary: 60F05, 62F35, 62G20

**Short title:** Partially linear single-index measurement error models.

*This research was supported in part by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse". The first author was supported by Alexander von Humboldt Foundation. The second author's was also partially supported by a grant from the National Cancer Institute (R29-CA-74552).

# 1 INTRODUCTION

We consider the semiparametric partially linear single index measurement error models (PLSIMEM),

$$Y_i = \eta(X_i^T\alpha_0) + Z_i^T\beta_0 + \varepsilon_i, \text{ with } \|\alpha_0\| = 1 \tag{1.1}$$

where $Y_i$ is a response variable; $X_i, Z_i$ are respectively exactly measured and error-prone covariates; $\eta(\cdot)$ is a smooth unknown function; and the errors $\varepsilon_i$ are independent with $E(\varepsilon_1|X,Z) = 0$ and $E(\varepsilon_1^2|X,Z) < \infty$. The restriction $\|\alpha_0\| = 1$ assures identifiability.

When $Z_i$ were exactly observed, this model is a natural generalization of the single-index and the partially linear models. The former case, where $\beta_0 = 0$, has been elaborately studied by Ichimura (1987), Härdle, Hall and Ichimura (1993) and more recently by Bonneu, Delecroix and Hristache (1995). The latter model with $X$ being a scalar and $\alpha = 1$ was introduced by Engle, et al. (1986) to study the effect of weather on electricity demand, and was further investigated by Heckman (1986), Chen (1988), Speckman (1988), Cuzick (1992a,b), Severini & Staniswalis (1994) and Mammen and Geer (1997).

Model (1.1) with exactly measured $Z$ is also a special case of the generalized partially linear single index models (GPLSIM) studied by Carroll et al (1997), where (1.1) is replaced by

$$g^{-1}\{E(Y_i, X_i, Z_i)\} = \eta(X_i^T\alpha_0) + Z_i^T\beta_0, \tag{1.2}$$

with $g$ being a known link function. The result in this paper proved some building blocks to investigate model (1.2) when $Z$ are measured with errors. Further detailed research is needed for the complete development of methodology in this general setting though.

We are interested in estimation of the unknown parameters $\beta_0$ and $\alpha$ and unknown function $\eta(\cdot)$ in model (1.1) when the covariates $Z$ are measured with error, and instead of observing $Z$, we observe its surrogate $W_i$. The description of PLSIMEM is completed by using an additive measurement error model to relate $W$ & Z:

$$W_i = Z_i + U_i, \tag{1.3}$$

where the measurement errors $U_i$ are independent, independent of $(Y_i, X_i, Z_i)$, and identically distributed symmetry random errors with covariance matrix $\Sigma_{uu}$. The linear and nonlinear measurement error literature has been surveyed by Fuller (1987) and Carroll,

Ruppert & Stephanski (1995), respectively. More recently, Liang, Härdle and Carroll (1999) consider a combination of a partially linear model and (1.3), which is a special case of PLSIMEM.

We briefly describe the motivation and the estimation procedure of Liang, et al. (1999), and extend it to a "pseudo-likelihood" type of procedure (pseudo-$\beta$ method) to accommodate the single index structure and the multivariate $X$. The estimation procedure and the asymptotic properties of this estimator is provided in section 2. To avoid the "curse of dimensionality" in nonparametric regression, a new estimator which utilizes local estimating equations is proposed in Section 3. A theoretical comparison the two proposed estimators is discussed in Section 4.

# 2 PSEUDO-$\beta$ METHOD

If the $Z$'s were completely observable, estimation of $\beta_0$ at ordinary rates of convergence can be obtained by the following algorithm. We assume that the observed data $(X_i, Z_i, Y_i), 1 \leq i \leq n$, are generated by the relation of (1.1), and $\eta(\cdot)$ is an unknown piecewise smooth and continuous univariate function. $\alpha_0$ is a $p-$variate unit vector. $E(\varepsilon_1|X,Z) = 0$ and $E(\varepsilon_1^2|X,Z) < \infty$. As indicated by Schick (1996),

$$E(Y|X_i) = \eta(X_i^T\alpha_0) + E(Z|X_i)^T\beta_0 \tag{2.1}$$

A combination of (1.1) and (2.1) yields that

$$Y_i - E(Y|X_i) = \{Z_i - E(Z|X_i)\}^T\beta_0 + \varepsilon_i$$

Suppose that the regression functions $E(Y|X)$ and $E(Z|X)$ are smooth, using the usual nonparametric smoothing techniques such as kernel regression, we can easily establish the estimator of $\beta_0$ by employing least squares, maximum likelihood or quasilikelihood methods. For surveys of various nonparametric methods, see Härdle (1990).

Let $\widehat{E}(Y|X)$ and $\widehat{E}(Z|X)$ be the kernel regressions of $Y$ and $Z$ on $X$ with bandwidth $h$ and kernel function $K_1(\cdot)$. Then

$$\widehat{\beta}_n^* = \Big[\sum_{i=1}^n \{Z_i - \widehat{E}(Z|X_i)\}\{Z_i - \widehat{E}(Z|X_i)\}^T\Big]^{-1} \sum_{i=1}^n \{Z_i - \widehat{E}(Z|X_i)\}\{Y_i - \widehat{E}(Y|X_i)\}. \tag{2.2}$$

Under some regularity conditions and a proper order of $h$,

$$n^{1/2}(\widehat{\beta}_n^* - \beta_0) \longrightarrow^{\mathcal{L}} \text{Normal}(0, \Gamma_{Z|X}^{-1}\Sigma_2\Gamma_{Z|X}^{-1}), \tag{2.3}$$

where $\Gamma_{Z|X}$ is the covariance matrix of $Z - E(Z|X)$ and $\Sigma_2$ is the covariance matrix of $\varepsilon\{Z - E(Z|X)\}$.

The least squares form of (2.2) can be used to show that if one ignores measurement error and replaces $Z$ by $W$, the resulting estimator is inconsistent for $\beta_0$. For the partially linear model, Liang, et al. (1999) apply the so-called "correction for attenuation" approach and propose to estimate $\beta$ by

$$\widehat{\beta}_n = \Big[\sum_{i=1}^{n}\{W_i - \widehat{E}(W|X_i)\}\{W_i - \widehat{E}(W|X_i)\}^T - n\Sigma_{uu}\Big]^{-1} \sum_{i=1}^{n}\{W_i - \widehat{E}(W|X_i)\}\{Y_i - \widehat{E}(Y|X_i)\}. \tag{2.4}$$

By deducting $n\Sigma_{uu}$ from the first term of the right-hand side of (2.4), they verify that $n^{-1}\sum_{i=1}^{n}\{W_i - \widehat{E}(W|X_i)\}\{W_i - \widehat{E}(W|X_i)\}^T$ and $n^{-1}\sum_{i=1}^{n}\{W_i - \widehat{E}(W|X_i)\}\{Y_i - \widehat{E}(Y|X_i)\}^T$ converge to $\Gamma_{Z|X} + \Sigma_{uu}$ and $\Gamma_{Z|X}\beta$, respectively, and prove that $\widehat{\beta}_n$ is consistent. The estimator $\widehat{\beta}_n$ can be directly adopted in the partially linear single-index setting by using a multivariate nonparametric estimator for $\widehat{E}(W|X_i)$. We shall show that $\widehat{\beta}_n$ is consistent as well as asymptotically normal. Before giving the first main result, we assume the following conditions.

**Condition 1.**

(i) $\Gamma_{Z|X} = E\{Z - E(Z|X))(Z - E(Z|X)\}^T$ is a positive-definite matrix.

(ii) Each entry of the Hessian matrices of $E(Z|X)$ and $E(Y|X)$ are continuous and squared integrable, where the $(i, j)$ entry of a Hessian matrix of $g(x)$ is defined as $\partial^2 g(x)\big/\partial x_i \partial x_j$.

(iii) $h_n \in [C_1 n^{-1/(p+4)}, C_2 n^{-1/(p+4)}]$ for $0 < C_1 < C_2$, where $p$ is the dimension of $X$.

(iv) $K_1(\bullet)$ is a bounded $p-$variate kernel function with compact support and a bounded Hessian.

$$\int K_1(u)du = 1 \text{ and } K_1(u) = K_1(-u)$$

(v) Weight functions $\omega_{ni}(\cdot)$ satisfy:

$$(i) \qquad \max_{1\le i\le n}\sum_{j=1}^{n}\omega_{ni}(T_j)=O(1),$$

$$(ii) \qquad \max_{1\le i,j\le n}\omega_{ni}(T_j)=O(b_n),$$

$$(iii) \qquad \max_{1\le i\le n}\sum_{j=1}^{n}\omega_{nj}(T_i)I(|T_j-T_i|>c_n)=O(c_n),$$

where $b_n=n^{-1+1/(p+4)}$, $c_n=n^{-1/(p+4)}\log n$.

**Theorem 2.1** *Suppose the condition 1 hold and $E(\varepsilon^4+\|U\|^4)<\infty$. Then $\widehat{\beta}_n$ is asymptotically normal;*

$$n^{1/2}(\widehat{\beta}_n-\beta_0)\to N(0,\Gamma_{Z|X}^{-1}\Sigma_{\beta P}\Gamma_{Z|X}^{-1}),$$

*where $\Sigma_{\beta P}=E[(\varepsilon-U^T\beta_0)\{Z-E(Z|X)\}]^{\otimes 2}+E\{(UU^T-\Sigma_{uu})\beta_0\}^{\otimes 2}+E(UU^T\varepsilon^2)$.*

*If $\varepsilon$ is homoscedastic and independent of $(Z,X)$, $\Sigma_{\beta P}$ can be simplified to $\sigma_*^2\Gamma_{Z|X}+\Sigma_M$, where $\sigma_*^2=E(\varepsilon-U^T\beta_0)^2$ and $\Sigma_M=E\{(UU^T-\Sigma_{uu})\beta_0\}^{\otimes 2}+\Sigma_{uu}\sigma^2$ with $A^{\otimes 2}=AA^T$.*

The proof of Theorem 2.1 is similar to that of Theorem 3.1 in Liang, et al. (1999), to which we refer for details. The key step is to obtain

$$\begin{aligned}\sqrt{n}(\widehat{\beta}_n-\beta_0)&=n^{-1/2}\Gamma_{Z|X}^{-1}\sum_{i=1}^{n}[\{Z_i+U_i-E(Z_i|X_i)\}(\varepsilon_i-U_i^T\beta_0)+\Sigma_{uu}\beta_0]+o_p(1)\\&=n^{-1/2}\Gamma_{Z|X}^{-1}\sum_{i=1}^{n}[\{Z_i-E(Z_i|X_i)\}(\varepsilon_i-U_i^T\beta_0)-(U_iU_i^T-\Sigma_{uu})\beta_0\\&\qquad +U_i\varepsilon_i]+o_p(1), \quad (2.5)\end{aligned}$$

which leads to the result of Theorem 2.1 directly.

Theorem 2.1 indicates that theoretically, when proper orders of bandwidths are chosen, the asymptotic distribution of the estimated coefficient, $\widehat{\beta}_n$, of $Z$ has the same structure regardless of the dimension of $X$. However, one should note that, in practice, if the dimension of $X$ is high, a large $n$ is required in order to reach the asymptotica. When $p$ is small, $\widehat{\beta}_n$ provides a simple consistent estimator. Without the assumption that $U$ is symmetry, one need to take into account the covariance between $U$ and $UU^T$. The exact asymptotic covariance in a more complicated form can be obtained using (2.5).

After obtaining the estimate of $\beta_0$, we pretend it were fixed and utilize the following "modified" model

$$Y_i - Z_i^T \widehat{\beta}_n = \eta(X_i^T \alpha_0) + \varepsilon_i \quad \text{and} \quad W_i = Z_i + U_i$$

to estimate $\alpha_0$ and $\eta(\cdot)$.

In literature, there exist several methods which estimate $\alpha_0$ at the $\sqrt{n}-$rate and $\eta(\cdot)$, at the usual nonparametric rate. For example, single-index estimation (Härdle, et al., 1993), projection pursuit regression (Friedman and Stuetzle. 1981) and Hall (1989), average derivative estimate (ADE) (Härdle and Stoker, 1989) and sliced inverse regression (Li, 1991). We estimate $\eta(\cdot)$ and $\alpha_0$ by a nonparametric kernel method. Suppose $\alpha$ is a unit $p-$vector and define

$$\eta(u|\alpha) = E(Y - Z^T \beta_0 | x^T \alpha = u).$$

A simple version of the estimate of $\eta(\cdot)$ suggests that

$$\widehat{\eta}(u|\alpha) = \Big\{\sum_{j=1}^{n} (Y_j - W_j^T \widehat{\beta}_n) K_{2h}(u - \alpha^T X_j)\Big\} \Big/ \Big\{\sum_{j=1}^{n} K_{2h}(u - \alpha^T X_j)\Big\}$$

where $K_{2h}(\cdot)$ is a one dimensional kernel function and $h$ is a corresponding bandwidth.

Let $\Lambda_i = X_i^T \alpha$ and $\widehat{\Lambda}_i = X_i^T \widehat{\alpha}$, an iterative estimation procedure can be described as follows.

Step 0. Obtain an initial estimate of $\widehat{\alpha}$.

Step 1. Find $\widehat{\eta}(u, \widehat{\alpha}, \widehat{\beta}_n) = \widehat{c}$ by minimizing

$$\sum_{i=1}^{n} \{c + W_i^T \widehat{\beta}_n - Y_i\}^2 K_{2h}(\widehat{\Lambda}_i - u)$$

Step 3. Update $\widehat{\alpha}$ by

$$\arg\min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} \{\widehat{\eta}(\Lambda_i, \widehat{\alpha}, \widehat{\beta}_n) + W_i^T \widehat{\beta}_n - Y_i\}^2$$

Iterate steps 1 and 2 until convergence.

Note that the estimate, $\widehat{\beta}_P$, is fixed as given in (2.4) throughout the iterations, as often been done for the pseudo-likelihood estimators in the parametric literature (Gong and

Sammaniego, 1981). We therefore call the foregoing estimators the "Pseudo-$\beta$" estimators and denote them by $\widehat{\alpha}_P$, $\widehat{\beta}_P$ and $\widehat{\eta}_P$, respectively. The sub-index "$P$" stands for "pseudo".

Before stating the asymptotic result about $\widehat{\alpha}_P$ and $\widehat{\eta}(u_0)$, we require the following condition.

**Condition 2.**

(i) The density function of $X$, $f(x)$, is bounded away from 0 and has two bounded derivative;

(ii) $\eta(\cdot)$ and the density function of $X^T\alpha_0$, $\gamma(\cdot)$, have two bounded, continuous derivatives;

(iii) $K_{2h}(\cdot)$ is supported on the interval $(-1, 1)$ and is a symmetric probability density, with a bounded derivative;

For simplicity in notation, we will denote $\mathcal{S}-E(\mathcal{S}|\Lambda)$ by $\widetilde{\mathcal{S}}$, for example, $\widetilde{X}_i = X_i - E(X|\Lambda_i)$ and $\widetilde{Z}_i = Z_i - E(Z|\Lambda_i)$.

**Theorem 2.2** *Under the conditions 1 and 2.*

$$\sqrt{n}\Gamma_{\alpha P}(\widehat{\alpha}_P - \alpha_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widetilde{X}_i\eta'(\Lambda_i)(\varepsilon_i - U_i^T\beta_0) - \Gamma_1(\widehat{\beta}_P - \beta_0) + o_p(1), \tag{2.6}$$

*where* $\Gamma_{\alpha P} = E\{\widetilde{X}\eta'(\Lambda)\}^{\otimes 2}$, $\Gamma_1 = E\{\widetilde{X}\widetilde{Z}^T\eta'(\Lambda)\}$ *with* $\Lambda = X^T\alpha_0$. *Furthermore* $\sqrt{n}(\widehat{\alpha}_P - \alpha_0)$ *is asymptotic* $N(0, \Gamma_{\alpha P}^{-1}\Sigma_{\alpha P}\Gamma_{\alpha P}^{-1})$. *When* $\varepsilon$ *is independent of* $X, Z$,

$$\Sigma_{\alpha P} = \Gamma_2\sigma_*^2 + \Gamma_1\Gamma_{Z|X}^{-1}\Sigma_M\Gamma_{Z|X}^{-1}\Gamma_1^T,$$

*where* $\Gamma_2 = E\{\widetilde{X}\eta'(\Lambda) - \Gamma_1\Gamma_{Z|X}^{-1}\widetilde{Z}\}^{\otimes 2} = \Gamma_{\alpha P} - \Gamma_1\Gamma_{Z|X}^{-1}\Gamma_1^T$.

Note that the second term in (2.6) corresponds to the extra variation due to estimating $\beta$. When $X$ and $Z$ are independent given $\Lambda$, $\Gamma_1 = 0$. That is, the asymptotic variance of $\widehat{\alpha}_P$ are the same regardless $\beta$ is estimated or not.

The outline of the proof is given at Appendix A.1, whereas the exact influence function of $\widehat{\alpha}_P$ is given at (A.7).

# 3 LOCAL QUASILIKELIHOOD METHOD

In section 2 we first directly derive an estimate of $\beta_0$, and then estimate $\alpha_0$ and $\eta(\cdot)$. Although this method is simple and intuitive, the estimator of $\beta_0$ does not fully utilize the information given in (1.1), but instead rely on a high dimension nonparametric estimation. It is therefore expected that a more efficient estimator can be derived. In this section, we consider a local estimating equation approach motivated by Severini and Staniswalis (1994) and Carroll, et al. (1997). The idea is that we can approximate $\eta(v)$, for $v$ in a neighborhood of $u$, by a constant or by a linear function:

$$\eta(v) \approx \eta(u) + \eta'(u)(v-u) \equiv a + b(v-u)$$

where $a = \eta(u)$ and $b = \eta'(u)$.

To estimate $\alpha_0$, $\beta_0$ and $\eta(\cdot)$, we first estimate $\eta(\cdot)$ as a function of $\alpha$ and $\beta$ to obtain $\widehat{\eta}(\alpha, \beta)$. Letting $\eta = \widehat{\eta}(\alpha, \beta)$, we then estimate the parametric component.

This procedure for the problem is equivalent to the following iterative algorithm without measured error (see Carroll, et al., 1997).

Step 0: Given initial values $(\widehat{\alpha}_1, \widehat{\beta}_L)$, set $\widehat{\alpha}_L = \widehat{\alpha}_1/\|\widehat{\alpha}_1\|$ and $\widehat{\Lambda}_i = X_i^T \widehat{\alpha}_L$.

Step 1: Find $\widehat{\eta}(u, \widehat{\alpha}_L, \widehat{\beta}_L) = \widehat{a}$ by maximizing the local log-quasilikelihood

$$\left.\begin{array}{ll} \text{local constant :} & \sum_{i=1}^n \{a + W_i^T \widehat{\beta}_L - Y_i\}^2 K_{3h}(\widehat{\Lambda}_i - u) \\ \text{local linear :} & \sum_{i=1}^n \{a + b(\widehat{\Lambda}_i - u) + W_i^T \widehat{\beta}_L - Y_i\}^2 K_{3h}(\widehat{\Lambda}_i - u) \end{array}\right\} \quad (3.1)$$

with respect to $a, b$.

Step 2: Update $(\widehat{\alpha}_L, \widehat{\beta}_L)$ by maximizing

$$\frac{1}{n}\sum_{i=1}^n \{\widehat{\eta}(\Lambda_i, \widehat{\alpha}_L, \widehat{\beta}_L) + W_i^T \beta - Y_i\}^2 - \beta^T \Sigma_{uu} \beta \quad (3.2)$$

with respect to $\alpha$ and $\beta$.

Step 3: Continue Steps 1 and 2 until convergence.

Step 4: Fix $(\alpha, \beta)$ at its estimated value from Step 3. The final estimate of $\eta(\cdot)$ is $\widehat{\eta}(u, \widehat{\alpha}_L, \widehat{\beta}_L) = \widehat{a}$ where $(\widehat{a}, \widehat{b})$ is obtained by (3.1).

We now concentrate on the local linear case and discuss the properties of $\widehat{\alpha}_L$ and $\widehat{\beta}_L$. In the proof of Theorem 3.1 we would point out that the asymptotic distributions of $\widehat{\alpha}_L$ and $\widehat{\beta}_L$ stay the same when $\widehat{\eta}$ is estimated using local constant smoother. We assume that $\widehat{\alpha}_L$ and $\widehat{\beta}_L$ are in a $\sqrt{n}-$ neighborhood of respectively $\alpha_0$ and $\beta_0$, i.e. $\widehat{\alpha}_L - \alpha_0 = O_P(n^{-1/2})$ and $\widehat{\beta}_L - \beta_0 = O_P(n^{-1/2})$.

**Theorem 3.1** *Under the conditions 1 and 2, and the assumption that the random vector $Z$ has a bounded support, for the estimator defined by (3.2), we obtain the following properties:*

$$\begin{aligned}\sqrt{n}\Gamma_{\alpha L}(\widehat{\alpha}_L - \alpha_0) = n^{-1/2}\sum_{i=1}^{n}[\{\widetilde{X}_i\eta'(\Lambda_i) - \Gamma_1\Gamma_{Z|\Lambda}^{-1}\widetilde{Z}_i\}(\varepsilon_i - U_i^T\beta_0) \\ +\Gamma_1\Gamma_{Z|\Lambda}^{-1}(U_iU_i^T - \Sigma_{uu})\beta_0 - \Gamma_1\Gamma_{Z|\Lambda}^{-1}U_i\varepsilon_i] + o_p(1)\end{aligned}$$

*and*

$$\begin{aligned}\sqrt{n}\Gamma_{\beta L}(\widehat{\beta}_L - \beta_0) = n^{-1/2}\sum_{i=1}^{n}[\{\widetilde{Z}_i - \Gamma_1^T\Gamma_{\alpha P}^{-1}\widetilde{X}_i\eta'(\Lambda_i)\}(\varepsilon_i - U_i^T\beta_0) \\ -(U_iU_i^T - \Sigma_{uu})\beta_0 + U_i\varepsilon_i] + o_p(1).\end{aligned}$$

*with $\Gamma_{Z|\Lambda} = E(\widetilde{Z}_i\widetilde{Z}_i^T)$; $\Gamma_{\alpha L} = \Gamma_{\alpha P} - \Gamma_1\Gamma_{Z|\Lambda}^{-1}\Gamma_1^T$, and $\Gamma_{\beta L} = \Gamma_{Z|\Lambda} - \Gamma_1^T\Gamma_{\alpha P}^{-1}\Gamma_1$. Therefore, $\sqrt{n}(\widehat{\alpha}_L - \alpha_0)$ is asymptotic normal with mean zero and asymptotic variance $\Gamma_{\alpha L}^{-1}\Sigma_{\alpha L}\Gamma_{\alpha L}^{-1}$, while the asymptotic distribution of $\sqrt{n}(\widehat{\beta}_L - \beta_0)$ is $N(0, \Gamma_{\beta L}^{-1}\Sigma_{\beta L}\Gamma_{\beta L}^{-1})$, where*

$$\Sigma_{\alpha L} = \Gamma_2^*\sigma_*^2 + \Gamma_1\Gamma_{Z|\Lambda}^{-1}\Sigma_M\Gamma_{Z|\Lambda}^{-1}\Gamma_1^T; \quad \Gamma_2^* = \Gamma_{\alpha P} - \Gamma_1\Gamma_{Z|\Lambda}^{-1}\Gamma_1^T,$$

*and*

$$\Sigma_{\beta L} = E\{\widetilde{Z} - \Gamma_1^T\Gamma_{\alpha P}^{-1}\widetilde{X}\eta'(\Lambda)\}^{\otimes 2}\sigma_*^2 + \Sigma_M,$$

*with $\sigma_*^2$ and $\Sigma_M$ defined in Theorem 2.1.*

# 4 COMPARISON AND DISCUSSION

Intuitively, the local likelihood method should gain efficiency comparing to the pseudo$-\beta$ method due to the dimension reduction. What we find is that the dimension reduction does contribute to the variance reduction, particularly when

$$\text{var}\{E(Z|X)\} \geq \text{var}\{E(Z|\Lambda)\}; \tag{4.3}$$

recall that $\Lambda_i = X_i^T\alpha$. Here, matrices $A \geq B$ implies $A - B$ is semi-positive-definite. (4.3) is generally true when the dimension of $X$ is high. Never the less, this is not sufficient to guarantee the superiority of the local likelihood method.

To appreciate this, we use the expression that

$$\text{avar}(\widehat{\beta}_P) = \sigma_*^2 \Gamma_{Z|X}^{-1} + \Gamma_{Z|X}^{-1} \Sigma_M \Gamma_{Z|X}^{-1}$$

and that

$$\text{avar}(\widehat{\beta}_L) = \sigma_*^2 {\Gamma^*_{Z|\Lambda}}^{-1} + {\Gamma^*_{Z|\Lambda}}^{-1} \Sigma_M {\Gamma^*_{Z|\Lambda}}^{-1},$$

where avar$(\cdot)$ indicates the asymptotic variance function, and $\Gamma^*_{Z|\Lambda} = \Gamma_{\beta L} = \Gamma_{Z|\Lambda} - \Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$. Note that $\Gamma_{Z|X} = E\{\text{var}(Z|X)\}$ and $\Gamma_{Z|\Lambda} = E\{\text{var}(Z|\Lambda)\}$. By the equality,

$$\text{var}(Z) = \text{var}\{E(Z|X)\} + E\{\text{var}(Z|X)\},$$

we have $\Gamma_{Z|X} \leq \Gamma_{Z|\Lambda}$ when (4.3) holds. Also note that comparing avar$(\widehat{\beta}_L)$ and avar$(\widehat{\beta}_P)$ is equivalent to comparing ${\Gamma^*_{Z|\Lambda}}^{-1}$ and $\Gamma_{Z|X}^{-1}$. However $\Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$ is semi-positive definite, which indicates a cost of estimating $\alpha$ in the local likelihood approach. This implies that avar$(\widehat{\beta}_P) \geq$ avar$(\widehat{\beta}_L)$ when $\Gamma_{Z|\Lambda} - \Gamma_{Z|X} \geq \Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$, that is, when the reduction in variation due to dimension reduction is larger than the cost of estimating $\alpha$ simultaneously.

Similarly, observing the asymptotic variance expressions of $\hat{\alpha}_P$ and $\hat{\alpha}_L$ in Theorem 2.2 and Theorem 3.1, respectively, and using similar arguments, it can be shown that

$$\Gamma_{Z|X}^{-1} \Sigma_M \Gamma_{Z|X}^{-1} > \Gamma_{Z|\Lambda}^{-1} \Sigma_M \Gamma_{Z|\Lambda}^{-1},$$

but $\Gamma_2^* > \Gamma_2$. That is, avar$(\widehat{\alpha}_P)$ could still be smaller than avar$(\widehat{\alpha}_L)$. When $X$ and $Z$ are independent, the two variances are exactly the same. Therefore, one may prefer the pseudo$-\beta$ method over the local quasilikelihood approach because of its simplicity when $X$ and $Z$ are weakly correlated and when the dimension of $X$ is low.

# 5 APPENDIX

## 5.1 APPENDIX A.1

**Lemma 5.1** *(Liang, 1999) Let $V_1, \ldots, V_n$ be independent random variables with means zero and finite $r-$th moment ($r \geq 2$), i.e., $\sup_{1 \leq j \leq n} E|V_j|^r \leq C < \infty$. Assume $(a_{ki}, k, i =$*

$1 \ldots, n)$ *be a sequence of positive numbers such that* $\sup_{1\le i,k\le n}|a_{ki}| \le n^{-p_1}$ *for some* $0 < p_1 \le 1$ *and* $\sum_{j=1}^{n} a_{ji} = O(n^{p_2})$ *for* $p_2 \ge \max(0, 2/r - p_1)$. *Then*

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

## 5.2 APPENDIX A.2: The proof of Theorem 2.2

The proofs of Theorems 2.2 and 3.1 follow a similar technique used to prove Theorem 4 of Carroll, et al. (1997). Therefore, only the key steps are given. Let $\Lambda_i = X_i^T\alpha_0$ and $\widehat{\Lambda}_i = X_i^T\widehat{\alpha}_P$. We will need the asymptotic expansions of $\widehat{\eta}(u_0, \widehat{\alpha}_P, \widehat{\beta}_P)$, which we state below and prove after the statement.

$$\widehat{\eta}(u_0, \widehat{\alpha}_P, \widehat{\beta}_P) - \eta(u_0) = \frac{1}{nf(u_0)}\sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)(\varepsilon_i - U_i^T\beta_0) - (\widehat{\beta}_P - \beta_0)^T E(Z|\Lambda) - (\widehat{\alpha}_P - \alpha_0)^T E\{X\eta'(\Lambda)|\Lambda\} + o_P(n^{1/2}). \quad (A.1)$$

Let $\widehat{c} = \widehat{\eta}(u_0, \widehat{\alpha}_P, \widehat{\beta}_P)$, which solves

$$0 = \frac{1}{n}\sum_{i=1}^{n} K_{2h}(\widehat{\Lambda}_i - u_0)\{Y_i - W_i^T\widehat{\beta}_P - \widehat{c}\}.$$

Via Taylor expansion and using the condition on $h$, we obtain

$$0 = \frac{1}{n}\sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)\{Y_i - W_i^T\beta_0 - c\} - B_{n1}(\widehat{c} - c) - (\widehat{\beta}_P - \beta_0)^T B_{n2} - (\widehat{\alpha}_P - \alpha_0)^T B_{n3} + o_P(n^{-1/2}) + O_P(h^2). \quad (A.2)$$

Here $B_{nj}$ $(j = 1, 2, 3)$ are the resulting sample matrices of kernel form. (A.1) is a direct result of (A.2) since $B_{n1} = 1 + o_p(1)$; $B_{n2} = E(Z|\Lambda = u_0)\{1 + o_p(1)\}$ and $B_{n3} = E(X\eta'(\Lambda)|\Lambda = u_0)\{1 + o_p(1)\}$.

We now prove Theorem 2.2. Some straight forward calculations show that $\widehat{\alpha}_P$ solves

$$\frac{1}{n}\sum_{i=1}^{n} X_i\eta'(\Lambda_i)[\varepsilon_i + \{\eta(\Lambda_i) - \widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_P, \widehat{\beta}_P)\} - W_i^T(\widehat{\beta}_P - \beta_0) - U_i^T\beta_0]\{1 + o_p(1)\} = 0. \quad (A.3)$$

By Taylor expansion and the continuity of $\eta'(\cdot)$, $\widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_P, \widehat{\beta}_P) - \eta(\Lambda_i)$ can be approximated by

$$\eta'(\Lambda_i)X_i^T(\widehat{\alpha}_P - \alpha_0) + \widehat{\eta}(\Lambda_i, \widehat{\alpha}_P, \widehat{\beta}_P) - \eta(\Lambda_i) + o_P(n^{-1/2}). \quad (A.4)$$

Substituting the expression for $\widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_P, \widehat{\beta}_P) - \eta(\Lambda_i)$ in (A.4) into (A.3), by further calculation and (A.1), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\varepsilon_i - X_i\eta'(\Lambda_i)\Big\{\varepsilon_i - X_i^T\eta'(\Lambda_i)(\widehat{\alpha}_P - \alpha_0) - \frac{1}{nf(\Lambda_i)}\sum_{j=1}^{n}K_{2h}(\Lambda_j - \Lambda_i)(\varepsilon_j - U_j^T\beta_0)$$
$$+E(Z^T|\Lambda_i)(\widehat{\beta}_P - \beta_0) + E(X^T\eta'(\Lambda)|\Lambda_i)(\widehat{\alpha}_P - \alpha_0)$$
$$-W_i^T(\widehat{\beta}_P - \beta_0) - U_i^T\beta_0\Big\} = o_p(n^{-1/2}),$$

which can be rewritten as

$$n^{-1/2}\sum_{i=1}^{n}X_i\eta'(\Lambda_i)(\varepsilon_i - U_i^T\beta_0) - n^{-1/2}\sum_{i=1}^{n}\frac{X_i\eta'(\Lambda_i)}{nf(\Lambda_i)}\sum_{j=1}^{n}K_{2h}(\Lambda_j - \Lambda_i)(\varepsilon_j - U_j^T\beta_0)$$
$$= n^{-1/2}\sum_{i=1}^{n}X_i\eta'(\Lambda_i)\begin{pmatrix}\widetilde{X}_i\eta'(\Lambda_i)\\ \widetilde{Z}_i + U_i\end{pmatrix}^T\begin{pmatrix}\widehat{\alpha}_P - \alpha_0\\ \widehat{\beta}_P - \beta_0\end{pmatrix} + o_p(1). \quad \text{(A.5)}$$

Note that the second term of the left-hand side of (A.5) is

$$n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i - U_i^T\beta_0)\frac{1}{n}\sum_{j=1}^{n}X_j\eta'(\Lambda_j)\frac{K_{2h}(\Lambda_j - \Lambda_i)}{f(\Lambda_j)},$$

which is shown to be

$$n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i - U_i^T\beta_0)E\{X\eta'(\Lambda)|\Lambda_i\} + o_p(1). \quad \text{(A.6)}$$

Combining (A.5) and (A.6), we obtain (2.6).

Recall that the asymptotic influence function of $\widehat{\beta}_P$ is given at (2.5), which leads to the following asymptotic expression of $\widehat{\alpha}_P$,

$$\sqrt{n}\Gamma_{\alpha P}(\widehat{\alpha}_P - \alpha_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\{\widetilde{X}_i\eta'(\Lambda_i) - \Gamma_1\Gamma_{Z|X}^{-1}\widetilde{Z}_i\}(\varepsilon_i - U_i^T\beta_0)$$
$$+\Gamma_1\Gamma_{Z|X}^{-1}\{(U_iU_i^T - \Sigma_{uu})\beta_0 - U_i\varepsilon_i\}\Big) + o_p(1), \quad \text{(A.7)}$$

Theorem 2.2 then follows the central limit theorem.

## 5.3 APPENDIX A.3: The proof of Theorem 3.1

As in A.2, we need the following asymptotic expansion of $\widehat{\eta}(u_0, \widehat{\alpha}_L, \widehat{\beta}_L)$ :

$$\widehat{\eta}(u_0, \widehat{\alpha}_L, \widehat{\beta}_L) - \eta(u_0) = \frac{1}{nf(u_0)}\sum_{i=1}^{n}K_{3h}(\Lambda_i - u_0)(\varepsilon_i - U_i^T\beta_0)$$
$$-(\widehat{\beta}_L - \beta_0)^TE(Z|\Lambda = u_0) - (\widehat{\alpha}_L - \alpha_0)^TE\{X\eta'(\cdot)|\Lambda = u_0\} + o_P(n^{1/2}), \quad \text{(A.8)}$$

which is analogous to (35) of Carroll, et al. (1997).

Let $a = \eta(u_0)$ and $b = h\eta'(u_0)$. The local linear estimates solve

$$0 = \frac{1}{n}\sum_{i=1}^{n} K_{3h}(\hat{\Lambda}_i - u_0) \begin{pmatrix} 1 \\ \hat{\Lambda}_i - u_0 \end{pmatrix} \{Y_i - W_i^T\hat{\beta}_L - \hat{a} - \hat{b}(\hat{\Lambda}_i - u_0)/h\},$$

which implies

$$\begin{aligned} 0 = \frac{1}{n}\sum_{i=1}^{n} K_{3h}(\Lambda_i - u_0) \begin{pmatrix} 1 \\ \Lambda_i - u_0 \end{pmatrix} \{Y_i - W_i^T\beta_0 - a - b(\Lambda - u_0)/h\} \\ -B_{n1} \begin{pmatrix} \hat{a} - a_0 \\ \hat{b} - b_0 \end{pmatrix} - (\hat{\beta}_L - \beta_0)^T B_{n2} - (\hat{\alpha}_L - \alpha_0)^T B_{n3} \\ +o_P(n^{-1/2}) + O_P(h^2). \end{aligned}$$

As in (A.5) $B_{nj}$ $(j = 1, 2, 3)$ are the resulting sample matrices of kernel form. Replacing $B_{nj}$ by their asymptotic counterparts, we obtain (A.8). Note that, when a local constant smoother is used, the proof of the expression at (A.8) follows the steps between (A.1) and (A.2) with $(\hat{\alpha}_P,\ \hat{\beta}_P)$ replaced by $(\hat{\alpha}_L,\ \hat{\beta}_L)$. A comparison between (A.1) and (A.8) indicates that whether one uses local constant or linear smoother, $\hat{\eta} - \eta$ has the same asymptotic expression. Since the rest of the proof only use this expression for $\hat{\eta}$, the resulting asymptotic distribution of $(\hat{\alpha}_L,\ \hat{\beta}_L)$ would be the same.

We know that $(\hat{\alpha}_L, \hat{\beta}_L)$ is the solution to

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \begin{Bmatrix} X_i\eta'(\Lambda_i) \\ W_i \end{Bmatrix} \Big[\varepsilon_i + \{\eta(\Lambda_i) - \hat{\eta}(\hat{\Lambda}_i, \hat{\alpha}_L, \hat{\beta}_L)\} \\ -W_i^T(\hat{\beta}_L - \beta_0) - U_i^T\beta_0\Big] + \begin{pmatrix} 0 \\ \Sigma_{uu}\hat{\beta}_L \end{pmatrix} = 0. \end{aligned} \tag{A.9}$$

Following (A.6) and (A.7) as well as using the expression for $\hat{\eta}(\hat{\Lambda}_i, \hat{\alpha}_L, \hat{\beta}_L) - \eta(\Lambda_i)$ in (A.9), we have that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} X_i\eta'(\Lambda_i) \\ Z_i + U_i \end{pmatrix} \Big\{\varepsilon_i - X_i^T\eta'(\Lambda_i)(\hat{\alpha}_L - \alpha_0) - \frac{1}{nf(\Lambda_i)}\sum_{j=1}^{n} K_{3h}(\Lambda_j - \Lambda_i) \\ (\varepsilon_j - U_j^T\beta_0) + E(Z^T|\Lambda_i)(\hat{\beta}_L - \beta_0) + E(X^T\eta'(\Lambda)|\Lambda_i)(\hat{\alpha}_L - \alpha_0) \\ +o_P(n^{-1/2}) - W_i^T(\hat{\beta}_L - \beta_0) - U_i^T\beta_0\Big\} + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{uu}\hat{\beta}_L \end{pmatrix} = 0, \end{aligned}$$

which implies

$$n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} X_i\eta'(\Lambda_i) \\ Z_i + U_i \end{pmatrix} (\varepsilon_i - U_i^T\beta_0) - n^{-1/2}\sum_{i=1}^{n} \begin{pmatrix} X_i\eta'(\Lambda_i) \\ Z_i + U_i \end{pmatrix} \frac{1}{nf(\Lambda_i)}$$

$$\times \left\{ \sum_{j=1}^{n} K_{3h}(\Lambda_j - \Lambda_i)(\varepsilon_j - U_j^T \beta_0) \right\}$$

$$= n^{1/2} \left[ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} X_i \eta'(\Lambda_i) \\ Z_i + U_i \end{pmatrix} \begin{pmatrix} \widetilde{X}_i \eta'(\Lambda_i) \\ \tilde{Z}_i + U_i \end{pmatrix}^T - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{uu} \end{pmatrix} \right] \begin{pmatrix} \widehat{\alpha}_L - \alpha_0 \\ \widehat{\beta}_L - \beta_0 \end{pmatrix}. \quad \text{(A.10)}$$

By interchanging the summations, the second term of the left-hand side is

$$n^{-1/2} \sum_{i=1}^{n} (\varepsilon_i - U_i^T \beta_0) \frac{1}{n} \sum_{j=1}^{n} \begin{pmatrix} X_j \eta'(\Lambda_j) \\ Z_j + U_j \end{pmatrix} \frac{K_{3h}(\Lambda_j - \Lambda_i)}{f(\Lambda_j)}.$$

This is essentially the same as

$$n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} E(X\eta'(\Lambda)|\Lambda_i) \\ E(Z|\Lambda_i) \end{pmatrix} (\varepsilon_i - U_i^T \beta_0). \quad \text{(A.11)}$$

Combining (A.10) and (A.11), we obtain:

$$n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} \{X_i - E(X_i|\Lambda_i)\}\eta'(\Lambda_i) \\ Z_i + U_i - E(Z_i|\Lambda_i) \end{pmatrix} (\varepsilon_i - U_i^T \beta_0) + \begin{pmatrix} 0 \\ \Sigma_{uu}\beta_0 \end{pmatrix}$$

$$= n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} X_i \widetilde{X}_i^T \eta'^2(\Lambda_i) & X_i(\tilde{Z}_i + U_i)^T \eta'(\Lambda_i) \\ (Z_i + U_i)\widetilde{X}_i^T \eta'(\Lambda_i) & (Z_i + U_i)(\tilde{Z}_i + U_i)^T - \Sigma_{uu} \end{pmatrix} \begin{pmatrix} \widehat{\alpha}_L - \alpha_0 \\ \widehat{\beta}_L - \beta_0 \end{pmatrix}.$$

Law of Large Numbers yields that

$$n^{1/2} \begin{pmatrix} \Gamma_{\alpha P} & \Gamma_1 \\ \Gamma_1^T & \Gamma_{Z|\Lambda} \end{pmatrix} \begin{pmatrix} \widehat{\alpha}_L - \alpha_0 \\ \widehat{\beta}_L - \beta_0 \end{pmatrix} = n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} \widetilde{X}_i \eta'(\Lambda_i)(\varepsilon_i - U_i^T \beta_0) \\ (\tilde{Z}_i + U_i)(\varepsilon_i - U_i^T \beta_0) + \Sigma_{uu}\beta_0 \end{pmatrix} + o_P(1), \quad \text{(A.12)}$$

which gives

$$n^{1/2}\Gamma_{\alpha P}(\widehat{\alpha}_L - \alpha_0) + n^{1/2}\Gamma_1(\widehat{\beta}_L - \beta_0) = n^{-1/2} \sum_{i=1}^{n} \widetilde{X}_i \eta'(\Lambda_i)(\varepsilon_i - U_i^T \beta_0) + o_p(1);$$

$$n^{1/2}\Gamma_1^T(\widehat{\alpha}_L - \alpha_0) + n^{1/2}\Gamma_{Z|\Lambda}(\widehat{\beta}_L - \beta_0) = n^{-1/2} \sum_{i=1}^{n} (\tilde{Z}_i + U_i)(\varepsilon_i - U_i^T \beta_0) + \Sigma_{uu}\beta_0 + o_P(1).$$

A direct simplification deduces the expressions of $n^{1/2}\Gamma_{\alpha P}(\widehat{\alpha}_L - \alpha_0) + n^{1/2}\Gamma_1(\widehat{\beta}_L - \beta_0)$ and $n^{1/2}\Gamma_1^T(\widehat{\alpha}_L - \alpha_0) + n^{1/2}\Gamma_{Z|\Lambda}(\widehat{\beta}_L - \beta_0)$. Their asymptotic distributions follow a central limit theorem. We thus complete the proof of Theorem 3.1.

## REFERENCES


Bonneu, M., Delecroix, M. and Hristache, M. (1995). *Semiparametric estimation of generalized linear models and related models.* Manuscript.

Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995), *Nonlinear Measurement Error Models.* Chapman and Hall, New York.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially single-index models. *Journal of the American Statistical Association*, **92**, 477-489.

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.

Gong, J. and Sammaniego, F.(1981). Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics*, **9**, 861-869.

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. and Gijbels, I. (1993). *Local Polynomial Modeling and Its Applications.* Monographs on Statistics and Applied Probability 66. Chapman and Hall.

Friedman, J. H. and Stuetzle,W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, **76**, 817-823.

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

Hall, P. (1989). On projection pursuit regression.*Annals of Statistics*, **17**, 573-588.

Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the methods of average derivatives. *Journal of the American Statistical Association*, **84**, 986-995.

Härdle, W. (1990). *Applied Nonparametric Regression,* Cambridge University Press, New York.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, **21**, 157-178.

Ichimura, H. (1987). *Estimation of Single Index Models.* Ph.D. dissertation, Department of Economics, MIT.

Heckman, N. E. (1986). Spline smoothing in partly linear models. *Journal of the Royal Statistical Society, Series B*, **48**, 244-248.

Li, K.C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, **86** 316-342.

Liang, H. (1999). An application of Bernstein inequality. *Econometric Theory*-problem and solution, **15**, 152.

Liang, H., Härdle, W. and Carroll, R.J. (1999). Large sample theory in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, in press.

Mammen, E. and van de Geer, S. (1997). Penalized estimation in partial linear models. *Annals of Statistics*, **25**, 1014-1035.

Schick, A. (1996). Weighted least squares estimates in partly linear regression models. *Statistics & Probability Letters*, **27**, 281-287.

Severini, T.A. and Staniswalis, J.G. (1994). Quasilikelihood estimation in semiparametric models, *Journal of the American Statistical Association*, **89**, 501-511.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B*, **50**, 413-436.