PARTIALLY LINEAR SINGLE-INDEX MEASUREMENT ERROR MODELS

Hua Liang and Naisyin Wang

St. Jude Children's Research Hospital and Texas A&M University

Abstract: We consider a partially linear single-index model $Y = \eta(Z^{\mathrm{T}}\alpha_0) + X^{\mathrm{T}}\beta_0 + \varepsilon$ when X is measured with additive error. Estimators in the literature are biased when the measurement errors are ignored. We propose two estimators in this setting and develop their asymptotic normality. We apply the proposed estimators to the analysis of dietary data, and provide the results of a simulation experiment to illustrate our approach.

Key words and phrases: Local linear regression, nonparametric regression, semi-parametric estimation.

1. Introduction

We consider the partially linear single-index model

$$Y = \eta(Z^{\mathrm{T}}\alpha_0) + X^{\mathrm{T}}\beta_0 + \varepsilon \text{ with } \|\alpha_0\| = 1, \tag{1.1}$$

where Y is response variable, $Z = (z_1, \ldots, z_p)^{\mathrm{T}}$ and $X = (x_1, \ldots, x_d)^{\mathrm{T}}$ are co-variates, α_0 and β_0 are parametric vectors to be estimated, $\eta(\cdot)$ is an unknown smooth function, $E(\varepsilon|Z, X) = 0$ and $E(\varepsilon^2|Z, X) < \infty$. The restriction $\|\alpha_0\| = 1$ assures identifiability.

Model (1.1) is a generalization of single-index models. Single-index models, $\beta_0 = 0$ in (1.1), have been studied in detail by Ichimura (1987) and Härdle, Hall and Ichimura (1993). When Z is a scalar quantity and $\alpha_0 = 1$, (1.1) is the partially linear model, introduced by Engle, Granger, Rice and Weiss (1986) to study the effect of weather on electricity demand, and was further investigated by Speckman (1988) and Severini and Staniswalis (1994). Härdle, Liang and Gao (2000) gave a comprehensive summary of statistical inference for partially linear models. A more general case of (1.1), generalized partially linear single index models, was studied by Carroll, Fan, Gijbels and Wand (1997), in which (1.1) is replaced by $g^{-1}\{E(Y|Z, X)\} = \eta(Z^{\mathrm{T}}\alpha_0) + X^{\mathrm{T}}\beta_0$, with g being a known link function.

In this paper, we are interested in estimating β_0, α_0 and the unknown function $\eta(\cdot)$ in (1.1) when the covariate X is measured with error-instead of observing

X, we observe its surrogate W. We assume an additive measurement error model to relate W and X:

$$W = X + U, \tag{1.2}$$

where the measurement error U, independent of (Y, Z, X), is a symmetric random error with a covariance matrix Σ_{uu}. We term (1.1) and (1.2) the partially linear single-index measurement error model (PLSMeM).

The literature on linear and nonlinear measurement error models has been reviewed by Fuller (1987) and by Carroll, Ruppert and Stefanski (1995). More recently, Liang, Härdle and Carroll (1999) considered a combination of the partially linear model and (1.2), which is a special case of PLSIMeM. In this paper, we focus on the situation that Σ_{uu} is known; the situation that Σ_{uu} is unknown can be dealt with by incorporating the estimated Σ_{uu} into the estimation procedures. Estimation of Σ_{uu} can be done based on replicates of W, as is discussed in, for example, Section 5 of Liang, Härdle and Carroll (1999).

We first use simple derivations to illustrate why the method proposed in Liang et al. (1999) can be easily extended to PLSIMeM. The only modification required is to replace an original univariate nonparametric regression with a multivariate one. The estimation procedure and the asymptotic properties of the estimator are provided in Section 2. To avoid the "curse of dimensionality" in nonparametric regression, a new estimator that uses local estimating equations is proposed in Section 3. Section 4 explores a theoretical comparison of the two classes of estimators. An analysis of data from the Women's Interview Survey of Health and a small simulation study are presented in Section 5. Concluding comments are given in Section 6. All proofs are given in the Appendix.

2. Pseudo-β Method

Assume that $(X_i, Z_i, Y_i, W_i, U_i)$ for $i = 1, \ldots, n$ are independent and identically distributed, (Z_i, X_i, Y_i) generated from (1.1), and W_i generated from (1.2). Note that $E(\varepsilon|Z, X) = 0$. It follows that $E(Y|Z_i) = \eta(Z_i^{\mathrm{T}}\alpha_0) + E(X|Z_i)^{\mathrm{T}}\beta_0$. This, along with (1.1), yields

$$Y_i - E(Y|Z_i) = \{X_i - E(X|Z_i)\}^{\mathrm{T}}\beta_0 + \varepsilon_i, \tag{2.1}$$

where the nonparametric term $\eta(Z_i^{\mathrm{T}}\alpha_0)$ is canceled from both sides of the original model. Note that based on the same principle, the partially linear model can also be maneuvered into (2.1) with a one dimensional Z. That is, when the regression functions $E(Y|Z)$ and $E(X|Z)$ are smooth and the X_i are completely observable, estimation of β_0 in both the partially linear and the partially linear single-index models can be obtained by the following algorithm.

Using the usual nonparametric smoothing methods, such as kernel regression, to approximate these two regression functions, a commonly used estimator is

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

where $\widehat{E}(Y|Z)$ and $\widehat{E}(X|Z)$ denote the kernel estimators of $E(Y|Z)$ and $E(X|Z)$.

Because the X_i are measured with error, it is well known that if one ignores measurement error and replaces X_i by W_i, the resulting estimator is inconsistent for β_0. Since (2.1) holds for the partially linear models and the partially linear single-index models, we note that the structure of (2.2) is identical to that of (3) in Liang et al. (1999). As a result, the *correction for attenuation* approach proposed by Liang et al. (1999) in the partially linear models remains applicable with a straightforward modification. More precisely, β_0, the parameter of interest in the PLSIMeM, is estimated by

$$\widehat{\beta}_{\mathrm{P}} = \Big[\sum_{i=1}^n \{W_i - \widehat{E}(W|Z_i)\}\{W_i - \widehat{E}(W|Z_i)\}^{\mathrm{T}} - n\Sigma_{uu}\Big]^{-1}$$
$$\sum_{i=1}^n \{W_i - \widehat{E}(W|Z_i)\}\{Y_i - \widehat{E}(Y|Z_i)\}. \qquad (2.3)$$

Here $\widehat{E}(\cdot|Z)$ denotes a p-dimensional kernel regression instead of the original 1-dimensional one in Liang et al. Deducting $n\Sigma_{uu}$ from the first term of the right-hand side of (2.3) and verifying that $n^{-1}\sum_{i=1}^n \{W_i - \widehat{E}(W|Z_i)\}\{W_i - \widehat{E}(W|Z_i)\}^{\mathrm{T}}$ and $n^{-1}\sum_{i=1}^n \{W_i - \widehat{E}(W|Z_i)\}\{Y_i - \widehat{E}(Y|Z_i)\}^{\mathrm{T}}$ converge to $\Gamma_{\mathrm{X|Z}} + \Sigma_{uu}$ and $\Gamma_{\mathrm{X|Z}}\beta$, respectively, we can prove that $\widehat{\beta}_{\mathrm{P}}$ is consistent. Here $\Gamma_{\mathrm{X|Z}}$ is the expectation of the covariance matrix of X given Z, and its exact expression is provided in condition 1 (i) below. We shall give the conditions that warrant this consistency as well as the asymptotic normality of $\widehat{\beta}_{\mathrm{P}}$.

Consider nonparametric estimates of $E(Y|Z = z)$ that are of a linear form, i.e., $\widehat{E}(Y|Z = z) = \sum_i w_{ni}(z)Y_i$ with $w_{ni}(z) = w_{ni}(z; Z_1, \ldots, Z_n)$. Common choices are kernel weights for which $w_{ni}(z)$ is defined as $w_{ni}(z) = n^{-1}K_H(Z_i - z)$, where $K_H(u) = |H|^{-1/2}K_1(H^{-1/2}u)$, $K_1(\cdot)$ is a bounded p-variate kernel with a compact support and a bounded Hessian matrix, $\int K_1(u)du = 1$, and H is a $p \times p$ symmetry positive definite bandwidth matrix depending on n; see, e.g., Ruppert and Wand (1994). For ease of presentation we assume here that H is a diagonal matrix, and take K_1 as the product of p symmetric univariate kernels.

We need the following.

Condition 1.

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

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

(iii) The diagonal elements of $H^{1/2}$ are of the same order $O(h_1)$, and $h_1 \in [C_1 n^{-1/(p+4)}, C_2 n^{-1/(p+4)}]$ for $0 < C_1 < C_2$, where p is the dimension of Z.

(iv) Weight functions $\omega_{ni}(\cdot)$ satisfy (a) $\max_{1 \le i \le n} \sum_{j=1}^{n} \omega_{ni}(Z_j) = O_{\mathrm{P}}(1)$; (b) $\max_{1 \le i,j \le n} \omega_{ni}(Z_j) = O_{\mathrm{P}}(b_n)$; and (c) $\max_{1 \le i \le n} \sum_{j=1}^{n} \omega_{nj}(Z_i) I(|Z_j - Z_i| > c_n) = O_{\mathrm{P}}(c_n)$, where $b_n = n^{-1+p/(p+4)}$, $c_n = n^{-1/(p+4)} \log n$, and $|Z_j - Z_i|$ denotes the Euclidean distance between the two vectors.

Condition 1 provides equivalent conditions to those behind Theorem 3.1 in Liang et al. (1999). In (iii), each univariate element in Z can have its own bandwidth, but all these p bandwidths share the same rate, as function of n and p. Condition 1 (iv) on the kernel weights is similar to Assumption 1.3 of Liang et al. (1999), but adapted to the current multivariate situation.

Theorem 2.1. *Suppose that Condition 1 holds and $E(\varepsilon^4 + \|U\|^4) < \infty$. Then $\widehat{\beta}_P$ is asymptotically normal:* $n^{1/2}(\widehat{\beta}_P - \beta_0) \to N(0, \Gamma_{X|Z}^{-1} \Sigma_{\beta P} \Gamma_{X|Z}^{-1})$, *where $\Sigma_{\beta P} = E[(\varepsilon - U^{\mathrm{T}}\beta_0)\{X - E(X|Z)\}]^{\otimes 2} + E\{(UU^{\mathrm{T}} - \Sigma_{uu})\beta_0\}^{\otimes 2} + E(UU^{\mathrm{T}}\varepsilon^2)$ and $A^{\otimes 2}$ denotes AA^{T}.*

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

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

$$\sqrt{n}(\widehat{\beta}_{\mathrm{P}} - \beta_0) = n^{-1/2} \Gamma_{X|Z}^{-1} \sum_{i=1}^{n} [\{X_i + U_i - E(X_i|Z_i)\}(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) + \Sigma_{uu}\beta_0] + o_{\mathrm{P}}(1)$$

$$= n^{-1/2} \Gamma_{X|Z}^{-1} \sum_{i=1}^{n} [\{X_i - E(X_i|Z_i)\}(\varepsilon_i - U_i^{\mathrm{T}}\beta_0)$$

$$- (U_i U_i^{\mathrm{T}} - \Sigma_{uu})\beta_0 + U_i \varepsilon_i] + o_{\mathrm{P}}(1), \tag{2.4}$$

which leads directly to the result.

Theorem 2.1 indicates that theoretically, when proper orders of bandwidths are chosen, the asymptotic distribution of the estimated coefficient, $\widehat{\beta}_{\mathrm{P}}$, of X

has the same structure regardless of the dimension of Z. However, one should note that, in practice, if the dimension of Z is high, a large n is required for the asymptotics to apply. When p is small, $\widehat{\beta}_{\mathrm{P}}$ provides a simple consistent estimator. Without the assumption that U is symmetric, one must take into account the covariance between U and UU^{T}. The exact asymptotic covariance in a more complicated form can be obtained by using (2.4).

After obtaining the estimate of β_0, we pretend it is fixed and use the following modified models $Y_i - X_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}} = \eta(Z_i^{\mathrm{T}}\alpha_0) + \varepsilon_i$ and $W_i = X_i + U_i$ to estimate α_0 and $\eta(\cdot)$. It is actually a single-index model.

In the literature, there are several methods to estimate α_0 at the $\sqrt{n}-$rate and $\eta(\cdot)$ at the usual nonparametric rate. For an example of single-index estimation see Härdle, Hall and Ichimura (1993), for projection pursuit regression see Friedman and Stuetzle (1981), for an average derivative estimate (ADE) see Härdle and Stoker (1989), and for sliced inverse regression see Li (1991).

We estimate $\eta(\cdot)$ and α_0 by a nonparametric kernel method. Suppose α is a unit p-vector and define $\eta(u, \alpha, \beta) = E(Y - X^{\mathrm{T}}\beta | Z^{\mathrm{T}}\alpha = u)$. Equality holds when replacing X by W in the right-hand-side. Because of this, the need for a measurement error correction arises when estimating β, but not when estimating η. The same phenomenon is observed in the estimation procedure proposed in the next section.

For estimating η, we use the local linear estimator investigated by Severini and Staniswalis (1994) and Carroll et al. (1997). The idea there is to approximate $\eta(v)$, for v in a neighborhood of u, by a linear function: $\eta(v) \approx \eta(u) + \eta'(u)(v - u) \equiv a + b(v - u)$, where $a = \eta(u)$ and $b = \eta'(u)$. Note that we can also use local linear estimates to replace the Nadaraya-Watson kernel regression estimates, $\widehat{E}(W|Z)$ and $\widehat{E}(Y|Z)$, in (2.3). The comparison, and equivalence, between these two type of estimators were studied in detail by Fan and Gijbels (1996). Theorem 2.1 still holds when local linear estimates are used. Putting $\Lambda_i = Z_i^{\mathrm{T}}\alpha$, the iterative estimation procedure can be described as follows.

Step 2.0. Obtain an initial value $\widehat{\alpha}_0$, for example by sliced inverse regression (Li, 1991) using $Y - W^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}}$ as responses, and set $\widehat{\alpha} = \widehat{\alpha}_0/\|\widehat{\alpha}_0\|$.

Step 2.1. Let $\widehat{\Lambda}_i = Z_i^{\mathrm{T}}\widehat{\alpha}$ and find $\widehat{\eta}(u, \widehat{\alpha}, \widehat{\beta}_{\mathrm{P}}) = \widehat{a}$ by minimizing

$$\sum_{i=1}^{n}\{a + b(\widehat{\Lambda}_i - u) + W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}} - Y_i\}^2 K_{2h}(\widehat{\Lambda}_i - u)$$

with respect to a and b. Here, $K_{2h}(\cdot) = 1/h_2 K_2(\cdot/h_2)$, $K_2(\cdot)$ is a one-dimensional kernel function and h_2 is the corresponding bandwidth.

Step 2.2. Update $\widehat{\alpha}$ by

$$\arg\min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} \{\widehat{\eta}(\Lambda_i, \alpha, \widehat{\beta}_{\mathrm{P}}) + W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}} - Y_i\}^2. \tag{2.5}$$

Iterate Steps 2.1 and 2.2 until convergence is achieved.

The estimate $\widehat{\beta}_{\mathrm{P}}$ is fixed in (2.3) throughout the iterations, as has often been done for the pseudo-likelihood estimators in the literature on parametric methods (Gong and Samaniego (1981)). We therefore call the foregoing estimators the pseudo-β estimators and denote them by $\widehat{\alpha}_{\mathrm{P}}$, $\widehat{\beta}_{\mathrm{P}}$ and $\widehat{\eta}_{\mathrm{P}}$, respectively ($P$ stands for pseudo).

We need the following.

Condition 2.
(i) The density function of Z, $f(z)$, is bounded away from 0 and has two bounded derivatives on its support.
(ii) $\eta(\cdot)$ and the density function of $Z^{\mathrm{T}}\alpha_0$, $\gamma(\cdot)$, have two bounded, continuous derivatives on their supports.
(iii) $K_2(\cdot)$ is supported on the interval $(-1, 1)$ and is a symmetric probability density, with a bounded derivative.

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

Theorem 2.2. *Under Conditions 1 and 2 and with $nh_2 \to \infty$ and $nh_2^4 \to 0$ as $n \to \infty$, $\sqrt{n}(\widehat{\alpha}_P - \alpha_0)$ converges in distribution to $N(0, \Gamma_{\alpha P}^{-1}\Sigma_{\alpha P}\Gamma_{\alpha P}^{-1})$, where $\Gamma_{\alpha P} = E\{\widetilde{Z}\eta'(\Lambda)\}^{\otimes 2}$ and $\Sigma_{\alpha P} = E\Big[\{\widetilde{Z}\eta'(\Lambda) - \Gamma_1\Gamma_{X|Z}^{-1}\widetilde{X}\}(\varepsilon - U^{\mathrm{T}}\beta_0) + \Gamma_1\Gamma_{X|Z}^{-1}\{(UU^{\mathrm{T}} - \Sigma_{uu})\beta_0 - U\varepsilon\}\Big]^{\otimes 2}$. When ε is independent of (Z, X), $\Sigma_{\alpha P} = \Gamma_2\sigma_*^2 + \Gamma_1\Gamma_{X|Z}^{-1}\Sigma_M\Gamma_{X|Z}^{-1}\Gamma_1^{\mathrm{T}}$, where $\Gamma_2 = E\{\widetilde{Z}\eta'(\Lambda) - \Gamma_1\Gamma_{X|Z}^{-1}\widetilde{X}\}^{\otimes 2} = \Gamma_{\alpha P} - \Gamma_1\Gamma_{X|Z}^{-1}\Gamma_1^{\mathrm{T}}$ and $\Gamma_1 = E\{\widetilde{Z}\widetilde{X}^{\mathrm{T}}\eta'(\Lambda)\}$ with $\Lambda = Z^{\mathrm{T}}\alpha_0$.*

When Z and X are independent given Λ, we have $\Gamma_1 = 0$ and $\Sigma_{\alpha P} = E\{\widetilde{Z}\eta'(\Lambda)(\varepsilon - U^{\mathrm{T}}\beta_0)\}^{\otimes 2}$. This suggests that the asymptotic variance of $\widehat{\alpha}_{\mathrm{P}}$ is the same regardless of whether β_0 is estimated or not. From (A.3) in the Appendix it can be seen that Γ_1 determines the extra variation due to estimation of the unknown β_0. When Γ_1 equals 0, this extra variation is zero.

The outline of the proof of Theorem 2.2 is given in the Appendix, with the exact influence function of $\widehat{\alpha}_{\mathrm{P}}$ given by (A.8). Due to space limitations, we focus on reporting the properties of the estimated α and β, but not of the estimated η. The reasons are two-fold: when estimating η, there is no need for a measurement error correction — the estimation of the parameters is the main interest; when α and β are $\sqrt{n}$ consistently estimated, the regular convergence rate of $\hat{\eta}$ can be

reached. The equivalent findings under different models have been investigated and reported by Carroll et al. (1997) and by Lin and Carroll (2001).

3. Modified Quasilikelihood Method

In Section 2 we first derived an estimate of β_0 directly, then estimated α_0 and $\eta(\cdot)$. Although this method is simple and intuitive, the estimator of β_0 does not fully use the information given in (1.1), but instead relies on a high dimension nonparametric estimation. It is therefore expected that a more efficient estimator can be derived. If $\eta(\cdot)$ were known and there were no measurement errors, the quasilikelihood estimator of (α_0, β_0) is the argument that minimizes

$$QL(\eta, \alpha, \beta, Y, Z, X) = \frac{1}{n}\sum_{i=1}^{n}\{\eta(Z_i^{\mathrm{T}}\alpha) + X_i^{\mathrm{T}}\beta - Y_i\}^2. \tag{3.1}$$

A quasilikelihood principle to account for measurement errors, as given in Chapter 7 of Carroll et al. (1995), is to focus on modeling the relationship between Y and (W, Z). This type of approach is different from the regression calibration idea in which X is replaced by the estimate of $\mathrm{E}(X|W, Z)$. When replacing X by W in (3.1), we note that (3.1) and $QL(\eta, \alpha, \beta, Y, Z, W) - \beta^{\mathrm{T}}\Sigma_{uu}\beta$ share the same asymptotic limit. That is, we can use the latter to replace the former as the objective function. We call this a modified quasilikelihood approach.

To estimate α_0, β_0 and $\eta(\cdot)$, we first estimate $\eta(\cdot)$ as a function of α and β, using a local linear smoother, and obtain $\widehat{\eta}(\cdot, \alpha, \beta)$. Letting $\eta(\cdot) = \widehat{\eta}(\cdot, \alpha, \beta)$, we then use the modified quasilikelihood idea to estimate the parametric component. The procedure may be described as the following iterative algorithm.

Step 3.0 Given initial values $(\widehat{\alpha}_1, \widehat{\beta}_{\mathrm{L}})$, set $\widehat{\alpha}_{\mathrm{L}} = \widehat{\alpha}_1/\|\widehat{\alpha}_1\|$ and $\widehat{\Lambda}_i = Z_i^{\mathrm{T}}\widehat{\alpha}_{\mathrm{L}}$.

Step 3.1. Find $\widehat{\eta}(u, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}}) = \widehat{a}$ by minimizing

$$\sum_{i=1}^{n}\{a + b(\widehat{\Lambda}_i - u) + W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{L}} - Y_i\}^2 K_{3h}(\widehat{\Lambda}_i - u) \tag{3.2}$$

with respect to a and b. As in Step 2.1, $K_{3h}(\cdot) = 1/h_3 K_3(\cdot/h_3)$.

Step 3.2. Update $(\widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}})$ by maximizing

$$\frac{1}{n}\sum_{i=1}^{n}\{\widehat{\eta}(Z_i^{\mathrm{T}}\alpha, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}}) + W_i^{\mathrm{T}}\beta - Y_i\}^2 - \beta^{\mathrm{T}}\Sigma_{uu}\beta \tag{3.3}$$

with respect to α and β.

Step 3.3. Iterate Steps 3.1 and 3.2 until convergence is achieved.

We select the initial estimates of $\widehat{\alpha}_{\mathrm{L}}$ and $\widehat{\beta}_{\mathrm{L}}$ to be consistent and have convergence rates faster than $\sqrt{nh_3} + h_3^2$, where h_3 is the bandwidth used in Step 3.1. This choice ensures that the asymptotic properties of $\widehat{\eta}$ obtained in Step 3.1 remain the same as those constructed with the true α_0 and β_0. This can be seen from evaluating the rates of convergence of the terms on the right hand side of (A.1). The estimates established in the previous section are $\sqrt{n}$ consistent and can serve as the initial estimates here.

Theorem 3.1. *Under Conditions 1 and 2 and with $nh_3 \to \infty$ and $nh_3^4 \to 0$ as $n \to \infty$, $\sqrt{n}(\widehat{\alpha}_L - \alpha_0)$ converges in distribution to $N(0, \Gamma_{\alpha L}^{-1}\Sigma_{\alpha L}\Gamma_{\alpha L}^{-1})$, and $\sqrt{n}(\widehat{\beta}_L - \beta_0)$ converges in distribution to $N(0, \Gamma_{\beta L}^{-1}\Sigma_{\beta L}\Gamma_{\beta L}^{-1})$, where $\Gamma_{X|\Lambda} = E(\widetilde{X}_i \widetilde{X}_i^{\mathrm{T}})$, $\Gamma_{\alpha L} = \Gamma_{\alpha P} - \Gamma_1 \Gamma_{X|\Lambda}^{-1}\Gamma_1^{\mathrm{T}}$, $\Gamma_{\beta L} = \Gamma_{X|\Lambda} - \Gamma_1^{\mathrm{T}}\Gamma_{\alpha P}^{-1}\Gamma_1$, $\Sigma_{\alpha L} = \Gamma_2^*\sigma_*^2 + \Gamma_1 \Gamma_{X|\Lambda}^{-1}\Sigma_M \Gamma_{X|\Lambda}^{-1}\Gamma_1^{\mathrm{T}}$, $\Gamma_2^* = \Gamma_{\alpha P} - \Gamma_1 \Gamma_{X|\Lambda}^{-1}\Gamma_1^{\mathrm{T}}$, and $\Sigma_{\beta L} = E\{\widetilde{X} - \Gamma_1^{\mathrm{T}}\Gamma_{\alpha P}^{-1}\widetilde{Z}\eta'(\Lambda)\}^{\otimes 2}\sigma_*^2 + \Sigma_M = \Gamma_{\beta L}\sigma_*^2 + \Sigma_M$, with σ_*^2 and Σ_M defined in Theorem 2.1.*

4. A Comparison of Two Estimators

In this section we compare the asymptotic variances of the two classes of estimators proposed in Sections 2 and 3. In general, this is difficult to do. Intuitively, the modified likelihood method should gain efficiency in comparison to the pseudo$-\beta$ method due to the reduction in dimension, but this is not always true. However, when ε is homoscedastic and independent of (Z, X), some insight is possible.

Let $\mathrm{avar}(\cdot)$ denote the asymptotic variance of an estimator. Then $\mathrm{avar}(\widehat{\beta}_{\mathrm{P}}) = \sigma_*^2\Gamma_{\mathrm{X|Z}}^{-1} + \Gamma_{\mathrm{X|Z}}^{-1}\Sigma_{\mathrm{M}}\Gamma_{\mathrm{X|Z}}^{-1}$ and $\mathrm{avar}(\widehat{\beta}_{\mathrm{L}}) = \sigma_*^2\Gamma_{\beta \mathrm{L}}^{-1} + \Gamma_{\beta \mathrm{L}}^{-1}\Sigma_{\mathrm{M}}\Gamma_{\beta \mathrm{L}}^{-1}$. We simplify the problem of comparing $\mathrm{avar}(\widehat{\beta}_{\mathrm{L}})$ and $\mathrm{avar}(\widehat{\beta}_{\mathrm{P}})$ to comparing $\Gamma_{\beta \mathrm{L}}^{-1}$ and $\Gamma_{\mathrm{X|Z}}^{-1}$.

Note that $\Gamma_{\mathrm{X|Z}} = E\{\mathrm{var}(X|Z)\}$, $\Gamma_{\mathrm{X|\Lambda}} = E\{\mathrm{var}(X|\Lambda)\}$, and that

$$\mathrm{var}(X) = \mathrm{var}\{E(X|Z)\} + \Gamma_{\mathrm{X|Z}} = \mathrm{var}\{E(X|\Lambda)\} + \Gamma_{\mathrm{X|\Lambda}},$$

$$\begin{aligned}\mathrm{var}\{E(X|Z)\} &= \mathrm{var}[E\{E(X|Z)|\Lambda\}] + E[\mathrm{var}\{E(X|Z)|\Lambda\}] \\ &= \mathrm{var}\{E(X|\Lambda)\} + E[\mathrm{var}\{E(X|Z)|\Lambda\}]. \end{aligned} \qquad (4.1)$$

The last equality is due to the fact that $\Lambda = Z^{\mathrm{T}}\alpha_0$ projects the space spanned by the original p dimensional vector Z onto a one-dimensional subspace. That is, with the second term in (4.1) being non-negative, we have

$$\mathrm{var}\{E(X|Z)\} \geq \mathrm{var}\{E(X|\Lambda)\}, \qquad (4.2)$$

where "$A \geq B$" means $A - B$ is semi-positive definite. When the dimension of Z is high, the difference can be large.

Now $\Gamma_{X|\Lambda} - \Gamma_{X|Z}$ is semi-positive definite because (4.2) holds. Recall that $\Gamma_{\beta L} = \Gamma_{X|\Lambda} - \Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$, and that $\Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$ is semi-positive definite, indication of the cost of estimating α in the local likelihood approach. This implies that $\mathrm{avar}(\widehat{\beta}_P) \geq \mathrm{avar}(\widehat{\beta}_L)$ when $\Gamma_{X|\Lambda} - \Gamma_{X|Z} \geq \Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1$, that is, when the reduction in variation due to the reduction in dimension is larger than the cost of estimating α simultaneously.

Recall that $\mathrm{avar}(\widehat{\alpha}_L) = \Gamma_{\alpha L}^{-1} \Sigma_{\alpha L} \Gamma_{\alpha L}^{-1}$ and $\mathrm{avar}(\widehat{\alpha}_P) = \Gamma_{\alpha P}^{-1} \Sigma_{\alpha P} \Gamma_{\alpha P}^{-1}$. Some straightforward but burdensome calculation gives $\Gamma_{\alpha L}^{-1} - \Gamma_{\alpha P}^{-1} = \Gamma_{\alpha P}^{-1} \Gamma_1 (\Gamma_{X|\Lambda} - \Gamma_1^T \Gamma_{\alpha P}^{-1} \Gamma_1) \Gamma_1^T \Gamma_{\alpha P}^{-1}$, and $\Sigma_{\alpha L} - \Sigma_{\alpha P} = \Gamma_1 (\Gamma_{X|Z}^{-1} - \Gamma_{X|\Lambda}^{-1}) \Gamma_1^T + \Gamma_1 (\Gamma_{X|\Lambda}^{-1} \Sigma_M \Gamma_{X|\Lambda}^{-1} - \Gamma_{X|Z}^{-1} \Sigma_M \Gamma_{X|Z}^{-1}) \Gamma_1^T$. We see that, even if $\Gamma_{X|Z}^{-1} > \Gamma_{X|\Lambda}^{-1}$, $\mathrm{avar}(\widehat{\alpha}_P)$ could still be smaller than $\mathrm{avar}(\widehat{\alpha}_L)$. It is difficult to say which of $\mathrm{avar}(\widehat{\alpha}_L)$ and $\mathrm{avar}(\widehat{\alpha}_P)$ is smaller.

When Z and X 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 Z and X are weakly correlated or when the dimension of Z is of low.

5. Numerical Examples

5.1. Simulation

We run a small simulation experiment with $n = 200$ and use data generated from the model

$$Y_i = X_i^T \beta_0 + \sin(Z_i^T \alpha_0) + \varepsilon_i \quad W_i = X_i + U_i,$$

where the errors ε_i are normally distributed with mean 0 and variance 0.2^2. The random vector X_i and the measurement error U_i are of dimension two; the distribution of U_i is $N(0, 0.15^2 I_2)$, with I_2 denoting the 2×2 identity matrix. We consider three cases: Case 1, in which Z_i is of dimension two, and X_i and Z_i are independent; Case 2, in which Z_i is also of dimension two but is correlated with X_i; and Case 3, in which Z_i is of dimension three and independent of X_i. In Cases 1 and 2, where Z_i is of dimension two, $\alpha_0 = (0.2, -0.7)/\sqrt{0.53}$. In Case 3, $\alpha_0 = (0.2, -0.7, 1)/\sqrt{1.53}$. For all cases, $\beta_0 = (0.8, 0.9)$. The distributions of X and Z in these three cases are described in detail below.

Case 1. The X_i are bivariate independent uniform $(0, 1)$ and the Z_i are bivariate $N(0, I_2)$; X and Z are independent.

Case 2. We generate X and Z from $N(0, V_{X,Z})$, where X and Z are of dimension two and $V_{X,Z} = \begin{pmatrix} 1.00 & -0.52 & 0.87 & 0.55 \\ -0.52 & 1.00 & -0.71 & -0.53 \\ 0.87 & -0.71 & 1.00 & 0.71 \\ 0.55 & -0.53 & 0.71 & 1.00 \end{pmatrix}$. We use this case to show the effect of moderate to strong correlation between Z and X on the variation of the two estimation procedures.

Case 3. The setup is identical to Case 1 except that the dimension of Z is now three. We use this to get a basic understanding of the effect of increasing dimensionality of Z.

In this experiment and in the following data analysis, three estimators are considered: the naive estimator obtained by ignoring measurement errors; pseudo $-\beta$; the modified quasilikelihood (MQL) estimates of the parameters α_0 and β_0. We use the quartic kernel $K(u) = 15/16(1 - u^2)^2 I_{(|u| \leq 1)}$. We generated 1,000 data sets in each of the three cases. The computation of the proposed estimates was conducted in XploRe — an advanced statistical environment developed by Härdle's team (http://www.xplore-stat.de/index_js.html). The estimated results are summarized in Table 1. Three estimates $\widehat{\eta}(\cdot)$ in Case 1, are given in Figure 1. The outcomes $\widehat{\eta}(\cdot)$ for Cases 2 and 3 are similar to $\widehat{\eta}(\cdot)$ in Case 1 and are not reported.

Table 1. The estimates with standard error (s.e.) of the parameters α_0 and β_0 obtained by three different methods for the simulated data.

Case	Parameter	True	Naive	Pseudo$-\beta$	MQL
1	β_1	0.8	0.65 (0.083)	0.73 (0.12)	0.75 (0.10)
	β_2	0.9	0.71 (0.081)	0.84 (0.09)	0.86 (0.09)
	α_1	0.275	0.315 (0.41)	0.281 (0.44)	0.278 (0.43)
	α_2	-0.962	-0.94 (0.38)	-0.95 (0.39)	-0.953 (0.42)
2	β_1	0.8	0.648 (0.125)	0.722 (0.42)	0.761 (0.24)
	β_2	0.9	0.724 (0.203)	0.831 (0.38)	0.867 (0.29)
	α_1	0.275	0.323 (0.43)	0.285 (0.53)	0.281 (0.41)
	α_2	-0.962	-0.925 (0.56)	-0.948 (0.48)	-0.952 (0.39)
3	β_1	0.8	0.649 (0.114)	0.726 (0.19)	0.754 (0.12)
	β_2	0.9	0.712 (0.142)	0.835 (0.13)	0.86 (0.16)
	α_1	0.162	0.134 (0.41)	0.181 (0.54)	0.178 (0.45)
	α_2	-0.566	-0.49 (0.36)	-0.42 (0.49)	-0.543 (0.40)
	α_3	0.808	0.69 (0.47)	0.76 (0.51)	0.78 (0.45)

The results correspond fairly well to our theory. As in the conventional measurement error models, the naive estimator of β_0 is biased and the estimates of $\eta(\cdot)$ are far from their true values. Both the pseudo$-\beta$ and the modified likelihood methods reduce the biases observed in the naive method. However, the variances obtained by using these two methods are larger than that of the naive estimator. In Case 1, the pseudo$-\beta$ and modified local likelihood estimators perform similarly and it is hard to decide which is preferable. In Case 2 where X and Z are correlated, and in Case 3 where the dimension of Z is 3, the variances of the pseudo$-\beta$ estimators are in general larger than those of the MQL estimators.
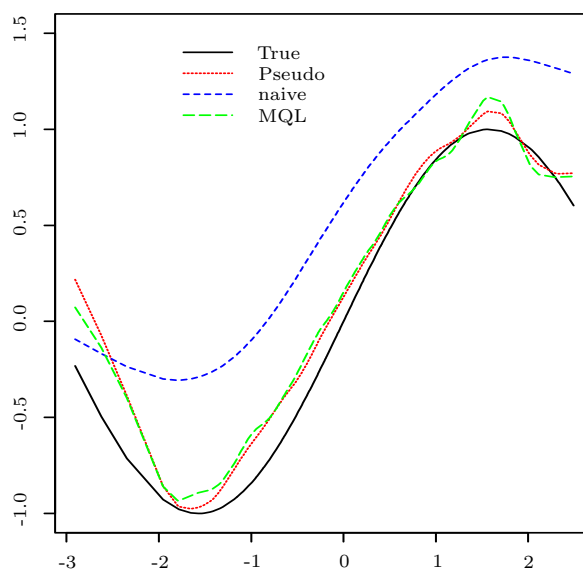
Figure 1. The true and estimated curves for the simulated data of Case 1. The dotted, short-dashed and long-dashed lines were obtained by using pseudo$-\beta$, MQL and naive methods. The true curve is indicated by the solid line.

5.2. Dietary data

The assessment of an individual's diet is difficult, but important in studying the relation between diet and cancer, and in monitoring dietary behavior. Several dietary instruments are used in nutrition research. Food Frequency Questionnaires (FFQ) are frequently administrated and from these, usual intake, Body Mass Index (BMI), and Age are recorded. Because of measurement errors and other sources of variability, one cannot observe true intake and other instruments, such as the 24-hour Food Recall or the multiple-day Food Record (both called FR), are used to obtain error-prone measurement of intake.

To illustrate the method, we consider data from the Women's Interview Survey of Health (WISH). We study a subset of the data consisting of 271 participants who completed an FFQ and six 24-hour food recalls on randomly selected days for at least two weeks. We want to analyze the relation between FFQ and usual intake. Experience indicates that FFQ depends nonlinearly on BMI and age. In our notation, $Z = (\text{BMI, age})^{\text{T}}$, X is the usual intake, measured with error, and Y is the FFQ. We have two replicates of W, the error prone measurement of the usual intake, and we use them to estimate the measurement error variance. The exact procedures, including the modified asymptotic variance for-

mulae for $\widehat{\alpha}$ and $\widehat{\beta}$, are described in Section 5 of Liang et al. (1999). The estimate for σ_{uu}^2 is $\widehat{\sigma}_{uu}^2 = \sum_{i=1}^n \sum_{j=1}^2 (W_{ij} - \bar{W}_i)^2/n$. The reliability ratio, σ_W^2/σ_X^2, is 1.35.

The estimated values of parameters of interest by using the naive, pseudo$-\beta$ and modified likelihood methods are presented in Table 2. The corresponding nonparametric estimates are provided in Figure 2.

Table 2. The estimates (s.e.) of the parameters α and β obtained by three different methods from the WISH data.

Parameter	naive	Pseudo$-\beta$	MQL
β (s.e)	0.397 (0.072)	0.527 (0.089)	0.531 (0.09)
α_1	0.362 (0.609)	0.387 (0.645)	0.383 (0.637)
α_2	0.086 (0.158)	0.142 (0.156)	0.138 (0.149)

Accounting for measurement errors, the estimates of β increase about 30%. The corresponding standard errors also increase by about 24%. The curve structures of the three estimated nonparametric functions are similar, with shifts reflecting the differences in $X^{\mathrm{T}}\widehat{\beta}$. We also fit a linear regression model to the dataset without considering measurement errors; the estimated mean function was FFQ $= 4.234 + 0.397\mathrm{FR} + 0.215\mathrm{BMI} + 0.076\mathrm{Age}$. Here, the estimated coefficient of FR is 0.397, and the positive coefficients of BMI and age suggest that FFQ increases with Age and BMI.
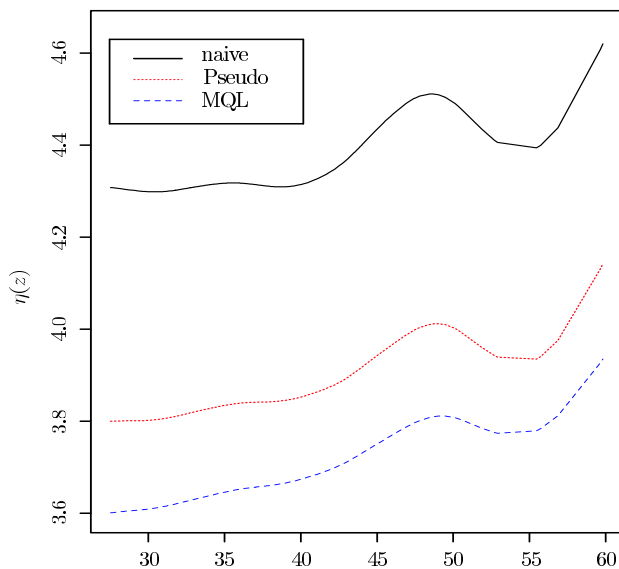

Figure 2. Estimated curves for the WISH data. The solid, dotted and dashed lines were obtained by using the naive, pseudo$-\beta$ and MQL methods.

6. Discussion

To handle measurement errors of parametric components in PLSIMeM, we have proposed two classes of estimators. One is based on a modified likelihood function; the other uses the method of moment (which we called the pseudo$-\beta$ method). The first method tends to give a more efficient estimator for the parametric component when the nonparametric component covariate, Z, is of high dimension. However, the implementation of this method requires the use of an iterative algorithm. The pseudo$-\beta$ estimator of the parametric component is easily understood and implemented without iteration. When the nonparametric and parametric components are independent, the two estimators are equivalent for large samples.

Fan and Troung (1993) examined the effect of measurement errors in nonparametric regression estimation. They proposed a class of kernel estimators based on deconvolution and showed that the optimal local and global rates of convergence of the kernel estimators were controlled by the tail behavior of the error distribution. In this paper we have not considered the case in which the nonparametric variables, Z, were measured with errors but the parametric variables, X, were measured exactly. This is a future research topic. Our present study has not considered models of general link functions, such as logit or probit, as in Carroll et al. (1997). Measurement error problems under this scenario is another topic that requires further study.

Appendix

The proofs of Theorems 2.2 and 3.1 follow a technique similar to that used by Carroll et al. (1997) to prove their Theorem 4. Only the key steps and departures from their procedure are given here. Details can be found in an earlier version of this article (Liang and Wang (2003)). Let $\Lambda_i = Z_i^{\mathrm{T}}\alpha_0$ and $\widehat{\Lambda}_i = Z_i^{\mathrm{T}}\widehat{\alpha}_{\mathrm{P}}$. We will need the asymptotic expansions of $\widehat{\eta}(u_0, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}})$, which we state below and prove after the statement.

$$\widehat{\eta}(u_0, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}}) - \eta(u_0) = \frac{1}{nf(u_0)} \sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)(\varepsilon_i - U_i^{\mathrm{T}}\beta_0)$$
$$- (\widehat{\beta}_{\mathrm{P}} - \beta_0)^{\mathrm{T}} E(X|\Lambda = u_0) - (\widehat{\alpha}_{\mathrm{P}} - \alpha_0)^{\mathrm{T}} E\{Z\eta'(\Lambda)|\Lambda = u_0\}$$
$$+ o_{\mathrm{P}}(n^{-1/2}) + O_{\mathrm{P}}(h_2^2). \tag{A.1}$$

Let $\widehat{a}$ and $\widehat{b}$ be the arguments that minimize the objective function given in Step 2.1 and $a = \eta(u_0)$ and $b = \eta'(u_0)$. The local linear estimates solve

$$0 = \frac{1}{n} \sum_{i=1}^{n} K_{2h}(\widehat{\Lambda}_i - u_0) \begin{pmatrix} 1 \\ \widehat{\Lambda}_i - u_0 \end{pmatrix} \{Y_i - W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}} - \widehat{a} - \widehat{b}(\widehat{\Lambda}_i - u_0)\}.$$

Focus on the top equation. $\widehat{a} = \widehat{\eta}(u_0, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}})$ solves $0 = 1/n \sum_{i=1}^{n} K_{2h}\{\widehat{\Lambda}_i - u_0)(Y_i - W_i^{\mathrm{T}}\widehat{\beta}_{\mathrm{P}} - \widehat{a} - \widehat{b}(\widehat{\Lambda}_i - u_0)\}$. Using Taylor expansion and eliminating higher order terms, we obtain

$$0 = \frac{1}{n} \sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)\{Y_i - W_i^{\mathrm{T}}\beta_0 - \eta(\Lambda_i)\} - B_{n1}\{\widehat{a} - \eta(u_0)\}$$
$$-(\widehat{\beta}_{\mathrm{P}} - \beta_0)^{\mathrm{T}} B_{n2} - (\widehat{\alpha}_{\mathrm{P}} - \alpha_0)^{\mathrm{T}} B_{n3} + o_{\mathrm{P}}(n^{-1/2}) + O_{\mathrm{P}}(h_2^2), \qquad (A.2)$$

where $B_{n1} = n^{-1} \sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)$, $B_{n2} = n^{-1} \sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)W_i$ and $B_{n3} = n^{-1} \sum_{i=1}^{n} K_{2h}(\Lambda_i - u_0)Z_i\eta'(u_0)$. Note that $B_{n1} = f(u_0) + o_{\mathrm{P}}(1)$. Divide all terms in (A.2) by $f(u_0)$ to get $B_{n2}/f(u_0) = E(X|\Lambda = u_0)\{1 + o_{\mathrm{P}}(1)\}$, and $B_{n3}/f(u_0) = E\{Z\eta'(u_0)|\Lambda = u_0\}\{1 + o_{\mathrm{P}}(1)\}$. We thus obtain equation (A.1).

Proof of Theorem 2.2. To prove Theorem 2.2, we first derive the following expression:

$$\sqrt{n}\Gamma_{\alpha\mathrm{P}}(\widehat{\alpha}_{\mathrm{P}} - \alpha_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{Z}_i\eta'(\Lambda_i)(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) - \sqrt{n}\Gamma_1(\widehat{\beta}_{\mathrm{P}} - \beta_0) + o_{\mathrm{P}}(1). \quad (A.3)$$

Recall that $\widehat{\alpha}_{\mathrm{P}}$ is the solution of (2.5). Taking a derivative of the objective function given in (2.5) on α, and some straightforward calculations, show that $\widehat{\alpha}_{\mathrm{P}}$ solves

$$\frac{1}{n} \sum_{i=1}^{n} Z_i\eta'(\Lambda_i)\Big[\varepsilon_i + \{\eta(\Lambda_i) - \widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}})\} - W_i^{\mathrm{T}}(\widehat{\beta}_{\mathrm{P}} - \beta_0) - U_i^{\mathrm{T}}\beta_0\Big]\{1 + o_{\mathrm{P}}(1)\} = 0.$$
$$(A.4)$$

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

$$\eta'(\Lambda_i)Z_i^{\mathrm{T}}(\widehat{\alpha}_{\mathrm{P}} - \alpha_0) + \widehat{\eta}(\Lambda_i, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}}) - \eta(\Lambda_i) + o_{\mathrm{P}}(n^{-1/2}). \qquad (A.5)$$

It follows from (A.5) and (A.4) that

$$\frac{1}{n} \sum_{i=1}^{n} Z_i\eta'(\Lambda_i)\Big[(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) + Z_i^{\mathrm{T}}\eta'(\Lambda_i)(\widehat{\alpha}_{\mathrm{P}} - \alpha_0) - W_i^{\mathrm{T}}(\widehat{\beta}_{\mathrm{P}} - \beta_0)$$
$$+ \widehat{\eta}(\Lambda_i, \widehat{\alpha}_{\mathrm{P}}, \widehat{\beta}_{\mathrm{P}}) - \eta(\Lambda_0)\Big]\{1 + o_{\mathrm{P}}(1)\} = 0.$$

This equation can be further expressed by using (A.1) and the conditions on h_2 as follows:

$$\frac{1}{n} \sum_{i=1}^{n} Z_i\eta'(\Lambda_i)\Big[\varepsilon_i - Z_i^{\mathrm{T}}\eta'(\Lambda_i)(\widehat{\alpha}_{\mathrm{P}} - \alpha_0) - \frac{1}{nf(\Lambda_i)} \sum_{j=1}^{n} K_{2h}(\Lambda_j - \Lambda_i)(\varepsilon_j - U_j^{\mathrm{T}}\beta_0)$$
$$+ E(X^{\mathrm{T}}|\Lambda_i)(\widehat{\beta}_{\mathrm{P}} - \beta_0) + E\{Z^{\mathrm{T}}\eta'(\Lambda)|\Lambda_i\}(\widehat{\alpha}_{\mathrm{P}} - \alpha_0)$$
$$- W_i^{\mathrm{T}}(\widehat{\beta}_{\mathrm{P}} - \beta_0) - U_i^{\mathrm{T}}\beta_0\Big] = o_{\mathrm{P}}(n^{-1/2}).$$

This may be written as

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

$$= n^{-1/2} \sum_{i=1}^{n} Z_i \eta'(\Lambda_i) \left\{ \begin{array}{c} \widetilde{Z}_i \eta'(\Lambda_i) \\ \widetilde{X}_i + U_i \end{array} \right\}^{\mathrm{T}} \left(\begin{array}{c} \widehat{\alpha}_{\mathrm{P}} - \alpha_0 \\ \widehat{\beta}_{\mathrm{P}} - \beta_0 \end{array} \right) + o_{\mathrm{P}}(1). \tag{A.6}$$

Note that the second term of the left-hand side of (A.6) is

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

Furthermore, we have

$$\frac{1}{n} \sum_{j=1}^{n} Z_j \eta'(\Lambda_j) \frac{K_{2h}(\Lambda_j - \Lambda_i)}{f(\Lambda_j)} = E\{Z\eta'(\Lambda)|\Lambda_i\} + o_{\mathrm{P}}(1).$$

The left-hand side of (A.6) is therefore shown to be

$$n^{-1/2} \sum_{i=1}^{n} (\varepsilon_i - U_i^{\mathrm{T}}\beta_0) \widetilde{Z}_i \eta'(\Lambda_i) + o_{\mathrm{P}}(1). \tag{A.7}$$

Replacing the left-hand side of (A.6) by (A.7) and noting that the terms in the right-hand side of (A.6) converge to Γ_1 and $\Gamma_{\alpha P}$, respectively, (A.3) follows.

From the asymptotic influence function of $\widehat{\beta}_{\mathrm{P}}$ given in (2.4), we deduce the asymptotic expression of $\widehat{\alpha}_{\mathrm{P}}$ as follows:

$$\sqrt{n}\Gamma_{\alpha P}(\widehat{\alpha}_{\mathrm{P}} - \alpha_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\{\widetilde{Z}_i \eta'(\Lambda_i) - \Gamma_1 \Gamma_{\mathrm{X|Z}}^{-1} \widetilde{X}_i\}(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) \right.$$

$$\left. + \Gamma_1 \Gamma_{\mathrm{X|Z}}^{-1}\{(U_i U_i^{\mathrm{T}} - \Sigma_{uu})\beta_0 - U_i \varepsilon_i\} \right] + o_{\mathrm{P}}(1). \tag{A.8}$$

The proof of Theorem 2.2 follows from the Central Limit Theorem.

Proof of Theorem 3.1. To complete the proof of Theorem 3.1, we derive

$$\sqrt{n}\Gamma_{\alpha L}(\widehat{\alpha}_{\mathrm{L}} - \alpha_0) = n^{-1/2} \sum_{i=1}^{n} \left[\{\widetilde{Z}_i \eta'(\Lambda_i) - \Gamma_1 \Gamma_{\mathrm{X|\Lambda}}^{-1} \widetilde{X}_i\}(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) \right.$$

$$\left. + \Gamma_1 \Gamma_{\mathrm{X|\Lambda}}^{-1}(U_i U_i^{\mathrm{T}} - \Sigma_{uu})\beta_0 - \Gamma_1 \Gamma_{\mathrm{X|\Lambda}}^{-1} U_i \varepsilon_i \right] + o_{\mathrm{P}}(1), \tag{A.9}$$

$$\sqrt{n}\Gamma_{\beta L}(\widehat{\beta}_{\mathrm{L}} - \beta_0) = n^{-1/2} \sum_{i=1}^{n} \left[\{\widetilde{X}_i - \Gamma_1^{\mathrm{T}} \Gamma_{\alpha P}^{-1} \widetilde{Z}_i \eta'(\Lambda_i)\}(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) \right.$$

$$\left. - (U_i U_i^{\mathrm{T}} - \Sigma_{uu})\beta_0 + U_i \varepsilon_i \right] + o_{\mathrm{P}}(1). \tag{A.10}$$

As in the proof of Theorem 2.2, we have the following asymptotic expansion of $\widehat{\eta}(u_0, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}})$:

$$
\begin{aligned}
\widehat{\eta}(u_0, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}}) - \eta(u_0) = {} & \frac{1}{nf(u_0)} \sum_{i=1}^{n} K_{3h}(\Lambda_i - u_0)(\varepsilon_i - U_i^{\mathrm{T}}\beta_0) \\
& - (\widehat{\beta}_{\mathrm{L}} - \beta_0)^{\mathrm{T}} E(X|\Lambda = u_0) - (\widehat{\alpha}_{\mathrm{L}} - \alpha_0)^{\mathrm{T}} E\{Z\eta'(\Lambda)|\Lambda = u_0\} \\
& + o_{\mathrm{P}}(n^{-1/2}) + O_{\mathrm{P}}(h_3^2);
\end{aligned} \tag{A.11}
$$

Note that (A.11) and (A.1) share the identical structure, as expected. We also note that for both equations, the second and the third terms on the right-hand side of the equation (A.11) go to zero faster than the first term, provided that the $\widehat{\alpha}$ and $\widehat{\beta}$ are consistent in a rate faster than $\sqrt{nh} + h^2$, where h is the bandwidth used in estimating η. This property has been used by Carroll et al. (1997) in their proofs implicitly. It follows from (3.3) that $(\widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}})$ satisfies

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \begin{Bmatrix} Z_i \eta'(\Lambda_i) \\ W_i \end{Bmatrix} \left[\varepsilon_i + \{\eta(\Lambda_i) - \widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}})\} - W_i^{\mathrm{T}}(\widehat{\beta}_{\mathrm{L}} - \beta_0) - U_i^{\mathrm{T}}\beta_0 \right] \\
& + \begin{pmatrix} 0 \\ \Sigma_{uu}\widehat{\beta}_{\mathrm{L}} \end{pmatrix} = o_p(1),
\end{aligned} \tag{A.12}
$$

where $\widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}})$ is given in Step 3.1. Following (A.7) and (A.8), as well as replacing $\widehat{\eta}(\widehat{\Lambda}_i, \widehat{\alpha}_{\mathrm{L}}, \widehat{\beta}_{\mathrm{L}}) - \eta(\Lambda_i)$ in (A.12) by (A.11), we have that

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \begin{Bmatrix} Z_i \eta'(\Lambda_i) \\ X_i + U_i \end{Bmatrix} \left[\varepsilon_i - Z_i^{\mathrm{T}}\eta'(\Lambda_i)(\widehat{\alpha}_{\mathrm{L}} - \alpha_0) - \frac{1}{nf(\Lambda_i)} \sum_{j=1}^{n} K_{3h}(\Lambda_j - \Lambda_i) \right. \\
& \quad (\varepsilon_j - U_j^{\mathrm{T}}\beta_0) + E(X^{\mathrm{T}}|\Lambda_i)(\widehat{\beta}_{\mathrm{L}} - \beta_0) + E\{Z^{\mathrm{T}}\eta'(\Lambda)|\Lambda_i\}(\widehat{\alpha}_{\mathrm{L}} - \alpha_0) \\
& \quad \left. + o_{\mathrm{P}}(n^{-1/2}) - W_i^{\mathrm{T}}(\widehat{\beta}_{\mathrm{L}} - \beta_0) - U_i^{\mathrm{T}}\beta_0 \right] + \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{uu}\widehat{\beta}_{\mathrm{L}} \end{pmatrix} = 0.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \begin{Bmatrix} Z_i \eta'(\Lambda_i) \\ X_i + U_i \end{Bmatrix} (\varepsilon_i - U_i^{\mathrm{T}}\beta_0) - n^{-1/2} \sum_{i=1}^{n} \begin{Bmatrix} Z_i \eta'(\Lambda_i) \\ X_i + U_i \end{Bmatrix} \frac{1}{nf(\Lambda_i)} \\
& \qquad \times \left\{ \sum_{j=1}^{n} K_{3h}(\Lambda_j - \Lambda_i)(\varepsilon_j - U_j^{\mathrm{T}}\beta_0) \right\} \\
& = n^{1/2} \left[\frac{1}{n} \sum_{i=1}^{n} \begin{Bmatrix} Z_i \eta'(\Lambda_i) \\ X_i + U_i \end{Bmatrix} \begin{Bmatrix} \widetilde{Z}_i \eta'(\Lambda_i) \\ \widetilde{X}_i + U_i \end{Bmatrix}^{\mathrm{T}} - \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{uu} \end{pmatrix} \right] \begin{pmatrix} \widehat{\alpha}_{\mathrm{L}} - \alpha_0 \\ \widehat{\beta}_{\mathrm{L}} - \beta_0 \end{pmatrix}. \tag{A.13}
\end{aligned}
$$

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

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

which equals

$$n^{-1/2} \sum_{i=1}^{n} \left[\begin{array}{c} E\{Z \eta'(\Lambda) | \Lambda_i\} \\ E(X | \Lambda_i) \end{array} \right] (\varepsilon_i - U_i^{\mathrm{T}} \beta_0) \{1 + O_{\mathrm{P}}(h_3^2) + o_{\mathrm{P}}(nh_3)^{-1/2}\}. \quad \text{(A.14)}$$

A combination of (A.13) and (A.14) yields

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

$$= n^{-1/2} \sum_{i=1}^{n} \left\{ \begin{array}{cc} Z_i \widetilde{Z}_i^{\mathrm{T}} \eta'^2(\Lambda_i) & Z_i (\widetilde{X}_i + U_i)^{\mathrm{T}} \eta'(\Lambda_i) \\ (X_i + U_i) \widetilde{Z}_i^{\mathrm{T}} \eta'(\Lambda_i) & (X_i + U_i)(\widetilde{X}_i + U_i)^{\mathrm{T}} - \Sigma_{uu} \end{array} \right\} \left(\begin{array}{c} \widehat{\alpha}_{\mathrm{L}} - \alpha_0 \\ \widehat{\beta}_{\mathrm{L}} - \beta_0 \end{array} \right).$$

A law of large numbers yields

$$n^{1/2} \left(\begin{array}{cc} \Gamma_{\alpha \mathrm{P}} & \Gamma_1 \\ \Gamma_1^{\mathrm{T}} & \Gamma_{\mathrm{X}|\Lambda} \end{array} \right) \left(\begin{array}{c} \widehat{\alpha}_{\mathrm{L}} - \alpha_0 \\ \widehat{\beta}_{\mathrm{L}} - \beta_0 \end{array} \right) = n^{-1/2} \sum_{i=1}^{n} \left\{ \begin{array}{c} \widetilde{Z}_i \eta'(\Lambda_i)(\varepsilon_i - U_i^{\mathrm{T}} \beta_0) \\ (\widetilde{X}_i + U_i)(\varepsilon_i - U_i^{\mathrm{T}} \beta_0) + \Sigma_{uu} \beta_0 \end{array} \right\}$$

$$+ o_{\mathrm{P}}(1). \quad \text{(A.15)}$$

It follows that

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

$$n^{1/2} \Gamma_1^{\mathrm{T}} (\widehat{\alpha}_{\mathrm{L}} - \alpha_0) + n^{1/2} \Gamma_{\mathrm{X}|\Lambda} (\widehat{\beta}_{\mathrm{L}} - \beta_0) = n^{-1/2} \sum_{i=1}^{n} (\widetilde{X}_i + U_i)(\varepsilon_i - U_i^{\mathrm{T}} \beta_0)$$

$$+ \Sigma_{uu} \beta_0 + o_{\mathrm{P}}(1).$$

A direct simplification deduces the expressions of (A.9) and (A.10). The asymptotic distributions follow from a Central Limit Theorem. The proof of Theorem 3.1 is complete.

Acknowledgement

References

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

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* **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. *J. Amer. Statist. Assoc.* **81**, 310-320.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman and Hall, London.

Fan, J. and Troung, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900-1925.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

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

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

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

Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models.* Springer Physica-Verlag, Heidelberg.

Ichimura, H. (1987). Estimation of single index models. Ph.D. dissertation, Department of Economics, MIT.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316-342.

Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Ann. Statist.* **27**, 1519-1535.

Liang, H. and Wang, N. (2003). Partially linear single-index measurement error models. Discussion paper, Department of Biostatistics, St Jude Children's Research Hospital, Memphis, U.S.A.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *J. Amer. Statist. Assoc.* **96**, 1045-1056.

Robinson, P. (1988). Root-n-consistent semiparametric regression. *Econometrica* **56**, 931-954.

Ruppert, D. and Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.* **22**, 1346-1370.

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.

Department of Biostatistics, St. Jude Children's Research Hospital, 332 North Lauderdale St., Memphis, TN 38105-2794, U.S.A.

E-mail: hua.liang@stjude.org

Department of Statistics, Texas A&M University, College Station TX 77843-3143, U.S.A.

E-mail: nwang@stat.tamu.edu