

Partially linear multivariate regression in the presence of measurement error

Seçil Yalaz^{1,a}, Müjgan Tez^b

^aDepartment of Statistics, Dicle University, Turkey;

^bDepartment of Statistics, Marmara University, Turkey

Abstract

In this paper, a partially linear multivariate model with error in the explanatory variable of the nonparametric part, and an m dimensional response variable is considered. Using the uniform consistency results found for the estimator of the nonparametric part, we derive an estimator of the parametric part. The dependence of the convergence rates on the errors distributions is examined and demonstrated that proposed estimator is asymptotically normal. In main results, both ordinary and super smooth error distributions are considered. Moreover, the derived estimators are applied to the economic behaviors of consumers. Our method handles contaminated data is founded more effectively than the semiparametric method ignores measurement errors.

Keywords: multivariate regression, partially linear model, errors in variables, kernel smoothing, asymptotic normality, Engel curves

1. Introduction

Let $(X_1, T_1), \dots, (X_n, T_n)$ be independent random variables in a model in which the conditional mean is linear in X_i but possibly nonlinear in $T_i \in \mathbb{R}^m$. Robinson (1988) considered a semiparametric partially linear model that relates a response Y to the predictors (X, T) through the function $X^T\beta + g(T)$. Applying the conditional expectation operator to this model, one obtains

$$\begin{aligned} E(Y_i|T_i) &= E(X_i^T\beta|T_i) + E(g(T_i)|T_i) + E(\varepsilon_i|T_i) \\ &= E(X_i|T_i)^T\beta + g(T_i), \end{aligned}$$

where ε_i are the model errors. Defining the conditional expectations as

$$\begin{aligned} g_y(t) &= E(Y_i|T_i = t), \\ g_x(t) &= E(X_i|T_i = t) \end{aligned}$$

the last expression can be written as

$$g_y(t) = g_x(t)^T\beta + g(t).$$

¹ Corresponding author: Department of Statistics, Dicle University, 21280 Sur, Diyarbakır, Turkey.
E-mail: secilyalaz@gmail.com

Assuming $m = 1$ and given $T_i = t$, the Nadaraya-Watson estimates of the conditional mean of Y_i and X_i take the following form:

$$g_y(t) = \frac{\sum_{i=1}^n Y_i K\left(\frac{t-T_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{t-T_i}{h_n}\right)},$$

$$g_x(t) = \frac{\sum_{i=1}^n X_i K\left(\frac{t-T_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{t-T_i}{h_n}\right)},$$

where K is the kernel function and h_n is the some sequence of bandwidths satisfying $h_n \rightarrow 0$ as $n \rightarrow \infty$. Subtracting these estimates from the original equation, one obtains

$$Y_i - g_y(t) = [X_i - g_x(t)]^T \beta + \varepsilon_i,$$

$$\tilde{Y}_i = \tilde{X}_i^T \beta + \varepsilon_i.$$

That is, the estimator of β can be predicted by the transformed equation

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

In this paper, we investigate the semiparametric regression problem with errors in the variables. This problem often occurs in econometrics when the sampling error stems from the deficiency of the measuring techniques.

Interest in studying and publishing the measurement error found in a semiparametric regression analysis is increasing. Some publications deal with errors in variable of parametric part, X (Liang *et al.*, 1999), errors in variable of nonparametric part, T (Liang, 2000) and errors in variables of both parametric and nonparametric parts (Zhu and Cui, 2003).

Suppose that the explanatory variable T is measured with error and $W_i = T_i + U_i$ is observed instead of T . Suppose also that the measurement errors U_i are i.i.d., and independent of (Y_i, X_i, T_i) . Using the kernel deconvolution method, Liang (2000) estimated the nonparametric regression function and the parameters of the parametric part in a semiparametric partially linear model. The author proposed

$$\omega_{ni}(t) = K_n\left(\frac{t - W_i}{h_n}\right) / \sum_j K_n\left(\frac{t - W_j}{h_n}\right)$$

with

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

where $\phi_K(s)$ is the Fourier transform of K and $\phi_U(s)$ is the characteristic function of the error variable U (Fan and Truong (1993) for details). Using this information, the author then defined

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

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

where $\tilde{Y}_i = Y_i - \sum_{j=1}^n \omega_{nj}(W_i) Y_j$ and $\tilde{X}_i = X_i - \sum_{j=1}^n \omega_{nj}(W_i) X_j$.

Toprak (2015) applied this model to measurements with errors following an unknown distribution. Using two repeated observations, the author estimated the parameters of the parametric part and determined the asymptotic normality properties of this estimator.

Some applications (such as financial applications) have focused on working with a multidimensional response variable. Models with multivariate nonparametric errors in the variables have been applied to dependent observations in Fan and Masry (1992), Masry (1991, 1993a, 1993b, 1993c). Moreover, Yalaz (2019) considered multivariate partially linear errors in variables regression while unknown function is multidimensional and measurement error is an $(n \times 1)$ matrix. However, models with semiparametric multivariate errors in the variables have not been clarified in the i.i.d case.

Now, we define a model with partially linear multivariate ($m \geq 1$) errors in the variables. For n observations, this model is expressed as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}^T \boldsymbol{\beta} + g(\mathbf{T}) + \boldsymbol{\varepsilon}, \\ \mathbf{W} &= \mathbf{T} + \mathbf{U}, \end{aligned} \quad (1.1)$$

where $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_m)$ is an $(n \times m)$ matrix of data, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ is an $(n \times p)$ matrix of variables, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)$ is a $(p \times m)$ matrix of unknown coefficients. g is unknown functions for $(n \times m)$ matrix of \mathbf{T} , $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_m) = (e_1, \dots, e_n)^T$ is an $(n \times m)$ matrix of random errors with mean vector 0 and a matrix of variances and covariances $\boldsymbol{\Sigma} = \sigma_{j_1 j_2}$ for $j_1, j_2 = 1, \dots, m$, and \mathbf{U} denotes the measurement errors. When the measurement errors \mathbf{U} are identically distributed and independent of $(\mathbf{Y}, \mathbf{X}, \mathbf{T})$, we observe \mathbf{W} instead of observing \mathbf{T} in model (1.1).

In the Section 2, we redefine the estimate of $\boldsymbol{\beta}$, seek a new estimator of $g(\mathbf{t})$, and perform the regression of \mathbf{Y} and \mathbf{X} on \mathbf{W} . Section 3 provides the main result. Income holds a primary position in the economy field. The income-expenditure behavior of consumers is among the most significant economic behaviors. The Engel curves explain the change of expenditure on different goods as a function of income (or total expenditure). Section 4 applies the proposed estimator to Engel curves. Our result is proven in the Appendix.

2. Construction of estimators

Let $\varphi(\cdot)$ be an integrable real-valued function defined on \mathbb{R}^m , $m \geq 1$. When $\boldsymbol{\beta}$ is known, absorbing $\mathbf{X}^T \boldsymbol{\beta}$ into \mathbf{Y} gives

$$\underbrace{\mathbf{Y} - \mathbf{X}^T \boldsymbol{\beta}}_{\mathbf{Y}^*} = g(\mathbf{t}) + \boldsymbol{\varepsilon}.$$

The more general regression function $g(\mathbf{t}) = E[\varphi(\mathbf{Y} - \mathbf{X}^T \boldsymbol{\beta}) | \mathbf{T} = \mathbf{t}] = E[\varphi(\mathbf{Y}^*) | \mathbf{T} = \mathbf{t}]$, $\mathbf{t} \in \mathbb{R}^m$ can be estimated similarly (Ioannides and Alevizos, 1997):

$$\hat{g}(\mathbf{t}) = \frac{\frac{1}{nh_n^m} \sum_{j=1}^n \varphi(Y_j^*) K_n\left[\frac{(\mathbf{t} - T_j)}{h_n}\right]}{\frac{1}{nh_n^m} \sum_{j=1}^n K_n\left[\frac{(\mathbf{t} - T_j)}{h_n}\right]}, \quad (2.1)$$

where $K_n(\mathbf{t}) = (1/(2\pi)^m) \int_{\mathbb{R}^m} \exp(-i\mathbf{s}\mathbf{t}) \{\phi_K(\mathbf{s})/\phi_U(\mathbf{s}/h_n)\} d\mathbf{s}$, $\mathbf{s}\mathbf{t} = s_1 t_1 + \dots + s_n t_n$.

To obtain $\hat{\boldsymbol{\beta}}$, we let $\boldsymbol{\omega}(\mathbf{t}) = (\omega_{n,h}(t_i, T_j))_{i,j}$ for $j = 1, \dots, n$, where $\omega_{n,h}(\mathbf{t}, T_j) = K_n[(\mathbf{t} - T_j)/h_n] / \sum_i K_n[(\mathbf{t} - T_i)/h_n]$, and obtain $\hat{g}(\mathbf{t}) = \sum_{j=1}^n \omega_{n,h}(\mathbf{t}, T_j) \varphi(Y_j^*)$. Substituting $\hat{g}(\mathbf{t})$ by $\hat{g}(\mathbf{w})$, the generalized

least squares estimator of β is found as

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_m) = (\tilde{X}^T \tilde{X})^{-1} (\tilde{X}^T \tilde{Y}), \quad (2.2)$$

where $\tilde{Y} = (I - \omega(\mathbf{w})) Y$ and $\tilde{X} = (I - \omega(\mathbf{w})) X$.

In the following section we examine the asymptotic properties of the estimator $\hat{\beta}$ using the rates of convergence L_∞ of the estimator \hat{g} .

3. Main results

We denote matrices by boldface letters and define $\gamma_j(\mathbf{t}) = E[x_{ij} | \text{vec}(\mathbf{T}) = \mathbf{t}]$. We then define $V_{ij} = x_{ij} - \gamma_j(\text{vec}(\mathbf{T}))$, where $\text{vec}(\mathbf{T})$ denotes the vector obtained by stacking the columns of \mathbf{T} , $i = 1, \dots, n$ and $j = 1, \dots, p$. In particular, for each $r \in \{1, \dots, m\}$, we have $(V_{ij})_r = x_{ij} - \gamma_j(T_{ir})$ and an $(n \times mp)$ matrix $\mathbf{V} = (\mathbf{V}_1, \dots, \mathbf{V}_m)$ for each $(n \times p)$ matrix of \mathbf{V}_r .

To construct the asymptotic normality, we impose Assumptions 1 and 2 below, and for the non-parametric estimation of g , we impose Assumptions 3–5 (Ioannides and Alevizos, 1997).

Assumption 1. Let $I = [0, 1]^m$, $m \geq 1$ be the unit cube. Then for a positive-definite \mathbf{B} matrix, we have $\sup_{\mathbf{t} \in I} E(\|X_1\|^3 | \mathbf{T} = \mathbf{t}) < \infty$ and $E(\mathbf{V}_1 \mathbf{V}_1^T) = \mathbf{B}$, where $\|X_1\|$ is the spectral norm of X_1 , defined as the square root of the maximum eigenvalue of $X_1^T X_1$.

Assumption 2. The numerator and denominator of $g(\mathbf{t}) = \int_{\mathbb{R}^m} \varphi(\mathbf{y}^*) f(\mathbf{t}, \mathbf{y}^*) d\mathbf{y}^* / f_n(\mathbf{t})$ and $\gamma_j(\mathbf{t})$ are Lipschitz of order θ for some $0 < \theta \leq 1$ on \mathbb{R}^m .

Assumption 3. The Kernel function $K(\mathbf{u})$ satisfies $\int_{\mathbb{R}^m} \|\mathbf{u}\|^\theta K(\mathbf{u}) d\mathbf{u} < \infty$, $0 < \theta \leq 1$.

Assumption 4.

- i. $\inf_{\mathbf{t} \in I} f_n(\mathbf{t}) > 0$, where I is given in Assumption 1.
- ii. The distribution of the error term is ordinary-smooth or super-smooth, respectively described as follows (Fan and Truong, 1993):
 - o A distribution is ordinary-smooth of order α if the characteristic function $\phi_\varepsilon(\mathbf{t})$ satisfies

$$d_0 \|\mathbf{t}\|^{-\alpha} \leq |\phi_\varepsilon(\mathbf{t})| \leq d_1 \|\mathbf{t}\|^{-\alpha}, \quad \text{as } \|\mathbf{t}\| \rightarrow \infty, \quad (3.1)$$

where d_0 , d_1 , and α are positive constants.

- o A distribution is super-smooth of order α if the characteristic function $\phi_\varepsilon(\mathbf{t})$ satisfies

$$d_0 \|\mathbf{t}\|^{\alpha_0} \exp\left(-\frac{\|\mathbf{t}\|^\alpha}{\zeta}\right) \leq |\phi_\varepsilon(\mathbf{t})| \leq d_1 \|\mathbf{t}\|^{\alpha_1} \exp\left(-\frac{\|\mathbf{t}\|^\alpha}{\zeta}\right), \quad \text{as } \|\mathbf{t}\| \rightarrow \infty, \quad (3.2)$$

where d_0 , d_1 , α , and ζ are positive constants and α_0 and α_1 are constants of any value.

Such ordinary-smooth distributions are gamma distributions of degree p or double exponential distributions, whereas super-smooth distributions are standard normal distributions when $\alpha = 2$ and Cauchy distributions when $\alpha = 1$.

Assumption 5.

i. An ordinary-smooth error distribution satisfies the following conditions:

1. $|\phi_\varepsilon(\mathbf{t})| > 0$ for all $\mathbf{t} \in \mathbb{R}^m$; $\|\mathbf{t}\|^\alpha |\phi_\varepsilon(\mathbf{t})| \geq D_\alpha$ for large $\|\mathbf{t}\|$, where $\alpha > 0$ and $D_\alpha > 0$.
2. $\int_{\mathbb{R}^m} \|\mathbf{t}\|^{2\alpha} |\phi_K(\mathbf{t})|^2 d\mathbf{t} < \infty$ and $\int_{\mathbb{R}^m} \|\mathbf{t}\|^\alpha \{\phi_K(\mathbf{t}) + \phi'_K(\mathbf{t})\} d\mathbf{t} < \infty$.

ii. A super-smooth error distribution satisfies the following conditions:

1. $|\phi_\varepsilon(\mathbf{t})| > 0$ for all $\mathbf{t} \in \mathbb{R}^m$; $|\phi_\varepsilon(\mathbf{t})| \geq d_0 \|\mathbf{t}\|^{\alpha_0} \exp(-\|\mathbf{t}\|^\alpha / \zeta)$ for large $\|\mathbf{t}\|$, where $d_0 > 0$, $\alpha > 0$ and α_0 is real.
2. $\phi_K(\mathbf{t}) = 0$ for $\|\mathbf{t}\| > d$ for some $d > 0$.

The following theorem states our main result, which is related to the limit distributions of $\hat{\boldsymbol{\beta}}$.

Theorem 1. Under Assumptions 1–5 and $E[\text{vec}(\boldsymbol{\varepsilon})^3 + \|\text{vec}(\mathbf{U})\|^3] < \infty$, the following hold:

- i. For an ordinary-smooth error distribution, we take $h_n = (\log n/n)^{\theta/2\theta+(2\alpha+1)m}$ for some $\theta > 0$ and $\alpha > 0$.
- ii. For a super-smooth error distribution, we take $h_n = (\log n)^{-\theta/\alpha}$ for some $\theta > 0$ and $\alpha > 0$.

Then

$$n^{\frac{1}{2}} [\text{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \Rightarrow N_{p \times m}(0, \boldsymbol{\Sigma} \otimes \mathbf{B}^{-1}) \quad \text{as } n \rightarrow \infty,$$

where $\Sigma_{ij} = \text{Cov}(\tilde{X}_i, \tilde{X}_j)$ and \mathbf{B} is given in Assumption 1.

Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1^T, \dots, \hat{\beta}_m^T)^T$ and $\boldsymbol{\beta} = (\beta_1^T, \dots, \beta_m^T)^T$. Then, for each $r \in \{1, \dots, m\}$, we further have

$$n^{\frac{1}{2}} (\hat{\beta}_r - \beta_r) \Rightarrow N_p(0, \sigma_{jj}^2 \mathbf{B}^{-1}) \quad \text{as } n \rightarrow \infty,$$

where $j = 1, \dots, m$.

4. Applications to Engel curves

Consumer behavior is how consumers make individual decisions about purchasing goods and services. Consumer behavior analysis provides significant benefits to business, marketing management, evaluation, and analysis of market opportunities. It is influenced by social, economic, psychological, and cultural factors, which underlie the characteristics of consumer behavior.

Income is an important component of consumption expenditure and consumer behavior. Income is the monetary expression of purchasing power, is acquired by a person or community during a period of work time and is periodically ongoing. Income actually dictates the power of individuals to purchase. Income distribution leads to needs, that influence income distribution among consumers in different spending groups. Consumption expenditure is the use of income on products and services that satisfy consumers' needs. Consumption is driven by compulsory needs as well as influenced by the close-acquaintance circles of societies, incomes, personal characteristics, and professions of individual consumers.

As confirmed in most studies, expenditure is highly affected by income level. The close relationship between consumer demand and income is often represented on Engel curves. Consumption

Table 1: Descriptive statistics for budget share data

Couples with one children								
Variable	Income class							
	(1)		(2)		(3)		(4)	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Food share	0.2254	0.0979	0.1882	0.0836	0.1735	0.0729	0.1051	0.0627
Housing share	0.2447	0.1412	0.2249	0.1285	0.2144	0.1132	0.2187	0.1105
Gasoline-motor oil share	0.0401	0.0534	0.0463	0.0672	0.0797	0.1336	0.0591	0.0893
Healthcare share	0.0790	0.0521	0.0712	0.0489	0.0675	0.0459	0.0536	0.0330
Telephone services share	0.0398	0.0262	0.0318	0.0206	0.0289	0.0204	0.0198	0.0137
Log total income	8.0133	0.2806	8.7182	0.1562	9.2625	0.1309	9.7236	0.2010
Sample size	33							

Couples with two children								
Variable	Income class							
	(1)		(2)		(3)		(4)	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Food share	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576
Housing share	0.2118	0.1091	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576
Gasoline-motor oil share	0.0703	0.0740	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576
Healthcare share	0.0618	0.0404	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576
Telephone services share	0.0291	0.0200	0.1544	0.0576	0.1544	0.0576	0.1544	0.0576
Log total income	8.0690	0.3180	8.7081	0.1256	9.2074	0.1122	9.7282	0.1787
Sample size	97							

patterns of households also respond to demographic characteristics; therefore, most empirical studies on Engel curves allow the parametric entry of demographic and other household characteristics that lead to semiparametric specification (Blundell *et al.*, 1998; Robinson, 1988).

In this section, we apply our approach to Engel curves to show the economic behavior of consumers. We selected data from the 2013 Consumer Expenditure Survey (CES) conducted by the U.S. Department of Labor (Bureau of Labor Statistics). The selected data were acquired from households with one or two children. We considered four income classes and five broad categories of goods: food, housing, gasoline and motor oil, healthcare, and telephone services. Table 1 gives the descriptive statistics of the sample used in the application study.

In the semiparametric approach, we specify partially linear multivariate for the budget shares and take the logarithm of the income for the four classes:

$$Y = X^T \beta + g(\ln T) + \varepsilon,$$

where $X^T \beta$ is a linear index of the observable exogenous regressors X and unknown parameters β . The income includes a normally distributed measurement error U :

$$W = T + U.$$

To choose the variance of U , we calculate the reliability ratio suggested by Fuller (1987):

$$r = \frac{\text{Var}(T)}{\text{Var}(T) + \sigma_u^2} = 0.85.$$

The parameters of the semiparametric model without measurement error (NoME) were estimated by the Nadaraya-Watson estimator with $h = 0.29$. The optimal bandwidth of the estimated semiparametric errors-in-variables (EIV) regression was determined by cross-validation (Stefanski and Carroll,

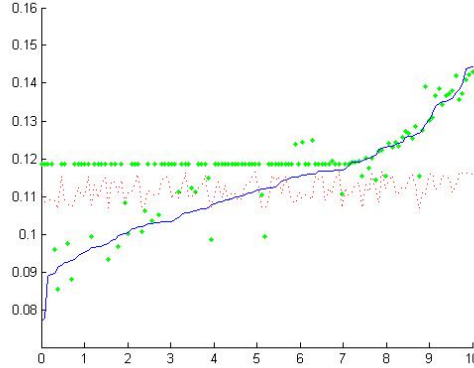


Figure 1: NoME fit and semiparametric EIV fit for the Engel curves.

1990). For the NoME and semiparametric EIV model, we selected a quartic kernel and a gaussian kernel, respectively. These kernels are respectively given by

$$\left(\frac{15}{16}\right)(1-u^2)^2 I(|u| \leq 1), \quad (4.1)$$

$$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right). \quad (4.2)$$

The deconvolution of the gaussian kernel is

$$K_n(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) \left[1 - \frac{0.125^2}{2h_n^2}(t^2 - 1)\right]. \quad (4.3)$$

Finally, we calculated the standard deviation of β and variance for each model. The standard deviation of β was $6,3468 \times 10^{-4}$ in the NoME case and $5,9412 \times 10^{-4}$ in the EIV case. The variance was $1,6959 \times 10^{-4}$ in the NoME case and $1,4860 \times 10^{-4}$ in the EIV case. The results are plotted in Figure 1. The data, NoME fit and EIV fit are represented by the solid line, dashed line and scatter points.

5. Discussion

In this paper, we extracted the partially linear multivariate regression function and its parameters from a model containing measurement errors. We assume that the parametric part \mathbf{X} of the model is measured exactly and the nonparametric part \mathbf{T} contains measurement errors. We also derived the asymptotic normality of the proposed estimator in the cases of ordinary-smooth and super-smooth distributions of the measurement error. Our method considers that the measurement error \mathbf{U} disturbs the statistics of $g_x(T)$ and $g_y(T)$. In the analysis on the economic behavior of consumers, our method handled contaminated data more effectively than the semiparametric method without measurement errors.

Appendix:

To prove the theorem, we need the following lemmas. The lemmas themselves are proven in Liang (2000).

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

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

Lemma 2. Suppose that Assumptions 1 and 3 hold. Then

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

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

Lemma 3. Suppose that Assumptions 1–5 hold. Then

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

where \mathbf{B} is given in Assumption 1.

Proof of Theorem 1:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{Y}} \\ &= (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T (\mathbf{I} - \omega(\mathbf{w})) (\mathbf{X}^T \boldsymbol{\beta} + g(\mathbf{t}) + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} [\tilde{\mathbf{X}}^T \tilde{\mathbf{g}} + \tilde{\mathbf{X}}^T (\mathbf{I} - \omega(\mathbf{w})) \boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta} + (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} [\tilde{\mathbf{X}}^T \tilde{\mathbf{g}} + \tilde{\mathbf{X}}^T \boldsymbol{\varepsilon} - \tilde{\mathbf{X}}^T \omega(\mathbf{w}) \boldsymbol{\varepsilon}]. \end{aligned}$$

We then have

$$\begin{aligned} \sqrt{n} \text{vec}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= n \left[I_m \otimes (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \right] \left\{ \frac{1}{\sqrt{n}} \left[I_m \otimes \tilde{\mathbf{X}}^T \right] \text{vec}(\tilde{\mathbf{g}}) + \frac{1}{\sqrt{n}} \left[I_m \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \omega(\mathbf{w})) \right] \text{vec}(\boldsymbol{\varepsilon}) \right\} \\ &= \mathbf{A} [\mathbf{C} + \mathbf{D}]. \end{aligned}$$

Consider the matrix $\mathbf{A} = n[I_m \otimes (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}]$. Lemma 3 means that \mathbf{A} converges to \mathbf{B}^{-1} . Consider the matrix $\mathbf{C} = (1/\sqrt{n})[I_m \otimes \tilde{\mathbf{X}}^T] \text{vec}(\tilde{\mathbf{g}})$.

$$\begin{aligned} \tilde{\mathbf{X}}^T \tilde{\mathbf{g}} &= [\tilde{\mathbf{V}} + \tilde{\boldsymbol{\gamma}}]^T \tilde{\mathbf{g}} \\ &= [(\mathbf{I} - \omega(\mathbf{w})) \mathbf{V} + \tilde{\boldsymbol{\gamma}}]^T \tilde{\mathbf{g}} \\ &= \mathbf{V}^T \tilde{\mathbf{g}} + \tilde{\boldsymbol{\gamma}}^T \tilde{\mathbf{g}} - \omega(\mathbf{w}) \mathbf{V}^T \tilde{\mathbf{g}} \\ &= \mathbf{F} + \mathbf{H} - \mathbf{J}. \end{aligned}$$

Taking $r = 2$, $V_k = V_{ij}$, $a_{ji} = \tilde{g}$, and $1/4 \leq p_1 \leq 1/3$, $p_2 = 1 - p_1$ in Lemma 1, the j^{th} column of the matrix $\mathbf{F} = \mathbf{V}^T \tilde{\mathbf{g}}$ becomes

$$\left| \sum_{i=1}^n V_{ij} \tilde{g} \right| = O\left(n^{-\frac{2p_1-1}{2}} \log n\right).$$

The j^{th} column of the matrix $\mathbf{H} = \tilde{\boldsymbol{\gamma}}^T \tilde{\boldsymbol{g}}$ is $\sum_{i=1}^n \tilde{\gamma}_{ij} \tilde{g}_i$. Using Abel's inequality $|\sum_{i=1}^n \tilde{\gamma}_{ij} \tilde{g}_i| \leq n \max_{i \leq n} |\tilde{g}_i| \max_{i \leq n} |\tilde{\gamma}_{ij}|$ and setting $G_j = g$ and $G_j = \gamma_j$ in Lemma 2, we get

$$\max_{i \leq n} |\tilde{g}_i| = \max_{1 \leq i \leq n} \left| g(T_i) - \sum_{k=1}^n \omega_{n,h}(W_i, W_k) g(T_k) \right| = o(1), \quad (\text{A.1})$$

and

$$\max_{i \leq n} |\tilde{\gamma}_j| = \max_{1 \leq i \leq n} \left| \gamma_j(T_i) - \sum_{k=1}^n \omega_{n,h}(W_i, W_k) \gamma_j(T_k) \right| = o(1), \quad (\text{A.2})$$

respectively. Therefore

$$\left| \sum_{i=1}^n \tilde{\gamma}_{ij} \tilde{g}_i \right| \leq o(1).$$

Using Abel's inequality, the j^{th} column of the matrix $\mathbf{J} = \boldsymbol{\omega}(\mathbf{w}) \mathbf{V}^T \tilde{\boldsymbol{g}}$ is given by

$$\left| \sum_{i=1}^n \sum_{k=1}^n \omega_{n,h}(W_i, W_k) V_{kj} \tilde{g}_i(T_i) \right| \leq n \max_{i \leq n} |\tilde{g}_i| \max_{i \leq n} \left| \sum_{k=1}^n \omega_{n,h}(W_i, W_k) V_{kj} \right|. \quad (\text{A.3})$$

Taking $r = 3$, $V_k = V_{kj}$, $a_{ji} = \omega_{n,h}(W_i, W_k)$, and $p_1 = 2/3$, $p_2 = 0$ in Lemma 1, we obtain

$$\max_{i \leq n} \left| \sum_{k=1}^n \omega_{n,h}(W_i, W_k) V_{kj} \right| = O(n^{-\frac{1}{3}} \log n). \quad (\text{A.4})$$

Using equations (A.3) and (A.4), we have

$$\left| \sum_{i=1}^n \sum_{k=1}^n \omega_{n,h}(W_i, W_k) V_{kj} \tilde{g}_i \right| \leq o(1).$$

By the above argument, the j^{th} column of the matrix $(1/\sqrt{n}) \tilde{\mathbf{X}}^T \tilde{\boldsymbol{g}}$ is given by $(1/\sqrt{n}) \sum_{i=1}^n \tilde{X}_i^T \tilde{g}_i = o(1)$.

Consider the matrix $\mathbf{D} = (1/\sqrt{n}) [I_m \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))] \text{vec}(\boldsymbol{\varepsilon})$. For a given arbitrary fixed vector $\mathbf{a} \in \mathbb{R}^{pm}/\{0\}$, let $\mathbf{c}^T = \mathbf{a}^T [I_m \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))]$, where $\mathbf{c}^T = (\mathbf{c}_1^T, \dots, \mathbf{c}_m^T)$ for each $\mathbf{c}_r^T = (c_{r1}, \dots, c_{rm})$.

We need to show that

$$\frac{1}{\sqrt{n}} \mathbf{a}^T [I_m \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))] \text{vec}(\boldsymbol{\varepsilon}) \xrightarrow{m} N_{p \times m} [0, \mathbf{a}^T (\boldsymbol{\Sigma} \otimes \mathbf{B}) \mathbf{a}]. \quad (\text{A.5})$$

Here we have

$$\mathbf{a}^T [I_m \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))] \text{vec}(\boldsymbol{\varepsilon}) = \mathbf{c}^T (\boldsymbol{\varepsilon}) = \sum_{r=1}^m \mathbf{c}_r^T \boldsymbol{\varepsilon}_r = \sum_{i=1}^n \mathbf{o}_i^T \mathbf{e}_i,$$

where $\mathbf{o}_i^T = (c_{1i}, \dots, c_{mi})$. Using the central limit theorem and assuming independent error vectors, we have $E(\mathbf{o}_i^T \mathbf{e}_i) = 0$ and $\text{Var}(\mathbf{o}_i^T \mathbf{e}_i) = \mathbf{o}_i^T \boldsymbol{\Sigma} \mathbf{o}_i$.

Hence,

$$\text{Var}\left(\sum_{i=1}^n \mathbf{o}_i^T \mathbf{e}_i\right) = \mathbf{a}^T \left[\boldsymbol{\Sigma} \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w})) (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} \right] \mathbf{a}.$$

Now let $n \rightarrow \infty$. Using Lemma 3, Equation (A.5) and Assumption 5, we have

$$n^{-1} \mathbf{a}^T \left[\boldsymbol{\Sigma} \otimes \tilde{\mathbf{X}}^T (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w})) (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} \right] \mathbf{a} \rightarrow \mathbf{a}^T [\boldsymbol{\Sigma} \otimes \mathbf{B}] \mathbf{a}.$$

From the Lindeberg condition, we need to show that

$$\max_{1 \leq i \leq n} c_{in}^2 \left/ \sum_{i=1}^n c_{in}^2 \right. \rightarrow 0 \quad (\text{A.6})$$

Equation (A.6) will hold if $\max_{1 \leq i \leq n} \text{Var}(\mathbf{o}_i^T \mathbf{e}_i) = o(n)$.

Let $b_r \in \mathbb{R}^m$ be a vector in which the r^{th} component is 1, and all other components are 0. Then

$$\begin{aligned} \max_{1 \leq i \leq n} |\mathbf{o}_i^T \boldsymbol{\Sigma} \mathbf{o}_i| &\leq \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq i \leq n} \sum_{r=1}^m c_{ri}^2 \leq \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq i \leq n} \left(m \max_{1 \leq r \leq m} c_{ri}^2 \right) \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left(\max_{1 \leq i \leq n} |c_{ri}| \right)^2 = m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (b_r^T \otimes I_n) \mathbf{c} \right\|_\infty^2 \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (b_r^T \otimes I_n) \left[I_m \otimes (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} \right] \mathbf{a} \right\|_\infty^2 \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (b_r^T \otimes (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}}) \mathbf{a} \right\|_\infty^2 \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (b_r^T \otimes (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}}) \mathbf{a} \right\|_\infty^2 \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} (b_r^T \otimes I_p) \mathbf{a} \right\|_\infty^2 \\ &= m \|\boldsymbol{\Sigma}\|_2 \max_{1 \leq r \leq m} \left\| (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} k_r \right\|_\infty^2, \end{aligned}$$

where $k_r = (b_r^T \otimes I_p) \mathbf{a} \in \mathbb{R}^p$.

$$\begin{aligned} \left\| (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \tilde{\mathbf{X}} k_r \right\|_\infty &= \left\| (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T \left[(\mathbf{I} - \boldsymbol{\omega}(\mathbf{w})) \mathbf{V} + \tilde{\boldsymbol{\gamma}} \right] k_r \right\|_\infty \\ &= \left\| (\mathbf{I} - \boldsymbol{\omega}(\mathbf{w}))^T (\mathbf{V} + \tilde{\boldsymbol{\gamma}} - \boldsymbol{\omega}(\mathbf{w}) \mathbf{V}) k_r \right\|_\infty \\ &\leq \left(1 + \left\| \boldsymbol{\omega}(\mathbf{w})^T \right\|_\infty \right) \left[\|\mathbf{V} k_r\|_\infty + \|\tilde{\boldsymbol{\gamma}} k_r\|_\infty + \|\boldsymbol{\omega}(\mathbf{w}) \mathbf{V} k_r\|_\infty \right]. \end{aligned}$$

From Assumption 5, $\|\boldsymbol{\omega}(\mathbf{w})^T\|_\infty$ are uniformly bounded. Assuming that $E|V_{ik}|^{2+\delta} \leq C < \infty$ for $\delta \geq 0$ for all i and k , using the Markov inequality

$$P\left(\max_{1 \leq i \leq n} |V_{ik}| \geq m_n\right) \leq \sum_i P(|V_{ik}| \geq m_n) \leq n C m_n^{-2-\delta}$$

for any constant m_n , and taking $m_n = n^{1/2} \log n$, we have

$$P(\|V_k\|_\infty > m_n) \rightarrow 0,$$

meaning that $\|V_k\|_\infty = o_p(n^{1/2})$ holds.

Using Equations (A.2) and (A.4), we respectively get $\|\tilde{\gamma}k_r\|_\infty = o(1)$ and $\|\omega(\mathbf{w})V_k\|_\infty = O(n^{-1/3} \log n)$.

Hence, by the Cramer-Wold device, we have

$$n^{-\frac{1}{2}} \left[I_m \otimes \tilde{X}^T (\mathbf{I} - \omega(\mathbf{w})) \right] \text{vec}(\boldsymbol{\varepsilon}) \xrightarrow{m} N_{p \times m} (0, \boldsymbol{\Sigma} \otimes \mathbf{B}).$$

□

References

- Blundell R, Duncan A, and Pendakur K (1998). Semiparametric estimation of consumer demand, *Journal of Applied Econometrics*, **13**, 435–461.
- Fan J and Masry E (1992). Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes, *Journal of Multivariate Analysis*, **43**, 237–271.
- Fan J and Truong YK (1993). Nonparametric regression with errors in variables, *The Annals of Statistics*, **21**, 1900–1925.
- Fuller WA (1987). *Measurement Error Models*, John Wiley & Sons, New York.
- Ioannides DA and Alevizos PD (1997). Nonparametric regression with errors in variables and applications, *Statistics and Probability Letters*, **32**, 35–43.
- Liang H (2000). Asymptotic normality of parametric part in partially linear model with measurement error in the non-parametric part, *Journal of Statistical Planning and Inference*, **86**, 51–62.
- Liang H, Hardle W, and Carroll RJ (1999). Estimation in a Semiparametric partially linear errors in variables model, *The Annals of Statistics*, **27**, 1519–1535.
- Masry E (1991). Multivariate probability density deconvolution for stationary random processes, *IEEE Transactions on Information Theory*, **37**, 1105–1115.
- Masry E (1993a). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes, *Journal of Multivariate Analysis*, **44**, 47–68.
- Masry E (1993b). Strong consistency and rates for deconvolution of multivariate densities of stationary processes, *Stochastic Processes and their Applications*, **47**, 53–74.
- Masry E (1993c). Multivariate regression estimation with errors in variables for stationary processes, *Journal of Nonparametric Statistics*, **3**, 13–36.
- Robinson PM (1988). Root-N-Consistent semiparametric regression, *Econometrica*, **56**, 931–954.
- Stefanski L and Carroll RJ (1990). Deconvoluting kernel density estimators, *Statistics*, **21**, 164–184.
- Toprak S (2015). *Semiparametric regression models with errors in variables*, (Doctoral dissertation), Dicle University, Turkey.
- Yalaz S (2019). Multivariate partially linear regression in the presence of measurement error, *ASIA Advances in Statistical Analysis*, **103**, 123–135.
- Zhu L and Cui H (2003). A semi-parametric regression model with errors in variables, *Scandinavian Journal of Statistics*, **30**, 429–442.

