WISE WORKING PAPER SERIES

WISEWP0614

Efficient Estimation of Partially Varying Coefficient Instrumental Variables Models

Zongwu Cai and Huaiyu Xiong

Octorber 1, 2006

厦门大学王亚南经济研究院
The Wang Yanan Institute for Studies in Economics, Xiamen University

Efficient Estimation of Partially Varying Coefficient Instrumental Variables Models*

Zongwu Cai[a,b] and Huaiyu Xiong[a]

[a]Department of Mathematics and Statistics,

University of North Carolina at Charlotte, Charlotte, NC 28223, USA

[b]Wang Yanan Institute for Studies in Economics, Xiamen University, China.

E-mail: zcai@uncc.edu

This version: October 1, 2006

We study a new class of semiparametric instrumental variables models with the structural function represented by a partially varying coefficient functional form. Under this representation, the models are linear in the endogenous/exogenous components with unknown constant or functional coefficients. As a result, the ill-posed inverse problem in a general nonparametric model with continuous endogenous variables does not exist under this setting. Efficient procedures are proposed to estimate both the constant and functional coefficients. Precisely, a three-step estimation procedure is proposed to estimate the constant parameters and the functional coefficients, we use the partial residuals and implement a nonparametric two-step estimation procedure. We establish the asymptotic properties for both estimators, including consistency and asymptotic normality. More importantly, it is also demonstrated that the constant parameters estimators are efficient, e.g., $\sqrt{n}$-consistent, and the functional coefficient estimators are oracle. A consistent estimation of the asymptotic covariance for both estimators is also provided. Finally, the high practical power of the resulting estimators is illustrated via both a Monte Carlo simulation study and an application to returns to education.

Keywords: Endogenous variables; Functional-coefficient models; Instrumental variables; Local linear fitting; Nonparametric smoothing; Simultaneous equations.

---

*Cai's research was supported, in part, by the National Science Foundation grant DMS-0404954 and funds provided by the University of North Carolina at Charlotte. We thank Xiaohong Chen for her constructive and useful comments and suggestions.

1 Introduction

Originated from supply and demand, measurement-in-error, and other problems in economics, structural models provide a useful framework for investigating causal relationships between two or more variables. A standard method to correctly account for the endogeneity represented in these problems is instrumental variables (IV) models. It is well known, however, that parametric IV models may be misspecified, and estimators obtained from misspecified models are often inconsistent. To deal with this issue, some nonparametric/semiparametric IV models has been proposed. For example, nonparametric IV models was first introduced by Newey and Powell (2003) to relax the strict limits imposed under parametric structural equations. Since then, nonparametric IV models have motivated many theoretical and methodological researches, and been found in many empirical applications[1]. Similar to standard nonparametric regression models, nonparametric IV models can permit more flexibility in the structure to exploit possible hidden patterns in the data. But when the number of covariates is large, they encounter the same "curse of dimensionality" problem.

To circumvent these difficulties, partial linear models proposed by Robinson (1988) and Andrews (1994) are brought into the IV settings. However, modifications have to be made to address the endogeneity in IV models. Previous work on semiparametric IV models include Pakes and Olley (1995), Park (2003), and Ai and Chen (2003), among others. For example, Pakes and Olley (1995) considered a semiparametric IV model with endogenous variables exhibited only in the parametric part. They proposed a generalized method of moment (GMM) type of method to estimate both the constant parameters and unknown functions and provided the consistency and asymptotic normality for both estimators. Further, Park (2003) extended the work to incorporate endogenous variables in both parametric and nonparametric parts, given that there exist two sets of instrumental variables satisfying an orthogonality condition, and then he obtained the similar asymptotic properties as in the previous model. Finally, Ai and Chen (2003) studied a more general semiparametric model with conditional moment restriction involving endogeneity and they mainly considered the efficient estimation and the $\sqrt{n}$ asymptotic normality result for the finite dimensional parameters but did not provide asymptotic distribution for the nonparametric components

[1] The recent developments include, for example, Newey, Powell and Vella (1999), Darolles, Florens, and Renault (2000), Ai and Chen (2003), Blundell and Powell (2003), Newey and Powell (2003), Carroll, Ruppert, Tosteson, Crainiceanu and Karagas (2004), Xiong (2004), Das (2005), Hall and Horowitz (2005), and Cai, Das, Xiong and Wu (2006).

because the exact leading bias term in series estimation is unknown. While we use the kernel method and derive the asymptotic normal distribution of our semiparametric estimator.

Having both "curse of dimensionality" and "ill-posed inverse" problems in mind and motivated by a real problem that investigates the empirical relationship between wages and education, we study a new class of the partially varying coefficient IV models, which are linear in the endogenous/exogenous components with either unknown functional coefficients of the predetermined variables or constant coefficients. On one hand, the partial linear structure reduces the dimension of the unknown functionals without losing the flexibility of the nonparametric functional form. On the other hand, it linearizes the nonparametric function in the endogenous components, so that the functional coefficients only depend on exogenous variables and the "ill-posed inverse" problem does not exhibit.

This paper has several contributions, described as follows. A fairly easy and yet efficient estimation procedure is proposed to estimate both the constant parameters and the coefficient functionals. The consistency and asymptotic normality for both constant and nonparametric parameter estimators are established. More importantly, it is shown that the estimator of constant parameters achieves the optimal parametric convergence rate at $\sqrt{n}$ and the estimator of functional coefficient is "oracle" in the sense that it has the same asymptotic property as if the constant parameters were known. Therefore, a partially varying coefficient IV model provides a handy tool for practitioners and applied researchers to handle real life data with endogeneity.

Our motivation on this research is from the following real problem. The interest is to investigate the empirical relationship between wages and education, using a random sample of young Australian female workers from the 1985 wave of the Australian Longitudinal Survey. It is well documented in the labor economics literature that the endogeneity of education in a wage model is due to unobservable heterogeneity in schooling choices; see, e.g., the review paper by Card (2001). In the same paper, he also suggested that if a wage model assumes the additive separability of education and experience, the returns to education is understated at higher levels of education because the marginal return to education is plausibly increasing in work experience. Observing our data, we also think a wage might be related other discrete variables like marital status, government employed, union status, and Australian-born that are more suitable for a parametric form. All these features suggest that a partially varying coefficient IV model would be a right target for this data set. The detailed analysis of this empirical example is reported in Section 5.

The rest of this paper is organized as follows. Section 2 presents the partially varying coefficient model and discusses its identification. Section 3 describes the three stage local linear regression estimators including an efficient estimate of constant parameters and functional coefficients in the model. The consistency and asymptotic normality of both estimators are given in Section 4, together with the consistent estimators of the asymptotic covariance matrices. Section 5 includes a simulation example and an application with real data to investigate the finite sample performance of our proposed estimation procedures. The proofs of our results are given in Section 6 with technical details relegated to the Appendix.

2 The Model

A partially varying coefficient IV model assumes the following form:

$$Y = g(\mathbf{X},\, \mathbf{Z}_1) + \varepsilon = \mathbf{g}_1(\mathbf{Z}_{11})^T\mathbf{Z}_{12} + \mathbf{g}_2(\mathbf{Z}_{11})^T\mathbf{X}_1 + \boldsymbol{\beta}_1^T\mathbf{Z}_{13} + \boldsymbol{\beta}_2^T\mathbf{X}_2 + \varepsilon, \tag{1}$$

where Y is an observable scalar random variable, $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ is a vector of endogenous variables including l-dimension vector $\mathbf{X}_1$ and p-dimension vector $\mathbf{X}_2$, $\mathbf{Z}_1 = (\mathbf{Z}_{11}^T, \mathbf{Z}_{12}^T, \mathbf{Z}_{13}^T)^T$ is a vector of exogenous variables, consisting of k-dimension vector $\mathbf{Z}_{11}$, d-dimension vector $\mathbf{Z}_{12}$ with its first element being one, and m-dimension vector $\mathbf{Z}_{13}$, $\boldsymbol{\mathcal{Z}} = (\mathbf{Z}_1^T, \mathbf{Z}_2^T)^T$ is a d_1-dimension vector with $\mathbf{Z}_2$ being a vector of q-dimension instrumental variables and $d_1 = k + d + m + q$, and $E(\varepsilon \,|\, \boldsymbol{\mathcal{Z}}) = 0$. The structural function $g(\cdot,\cdot)$ includes varying coefficients $\mathbf{g}_1(\cdot) = (g_1(\cdot), \ldots, g_d(\cdot))^T$ and $\mathbf{g}_2(\cdot) = (g_{d+1}(\cdot), \ldots, g_{d+l}(\cdot))^T$, and constant coefficients $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_m)^T$ and $\boldsymbol{\beta}_2 = (\beta_{m+1}, \ldots, \beta_{m+p})^T$. Like the model studied by Park (2003), the endogenous variables are allowed in both parametric and nonparametric parts. Note that if there is not any endogenous variable at all, model (1) becomes the partially varying coefficient regression model studied by Zhang, Lee and Song (2002) and Ahmad, Leelahanon and Li (2005) and it reduces to the model in Cai, Das, Xiong and Wu (hereafter, CDXW, 2006) if there is not a parametric part. Further, if $\mathbf{X}$ is a discrete endogenous variable, then model (1) covers the model studied by Das (2005) as a special case. Therefore, model (1) is a very general model.

Taking expectation on both sides of the structural equation (1), conditioning on $\boldsymbol{\mathcal{Z}}$, we obtain the following reduced form for Y:

$$E(Y|\boldsymbol{\mathcal{Z}}) = \mathbf{g}_1(\mathbf{Z}_{11})^T \mathbf{Z}_{12} + \mathbf{g}_2(\mathbf{Z}_{11})^T \boldsymbol{\pi}_1(\boldsymbol{\mathcal{Z}}) + \boldsymbol{\beta}_1^T \mathbf{Z}_{13} + \boldsymbol{\beta}_2^T \boldsymbol{\pi}_2(\boldsymbol{\mathcal{Z}}) \tag{2}$$

with $\pi_j = \pi_j(\boldsymbol{\mathcal{Z}}) = E(X_j|\boldsymbol{\mathcal{Z}})$, $\boldsymbol{\pi}_1(\boldsymbol{\mathcal{Z}}) = E(\mathbf{X}_1|\boldsymbol{\mathcal{Z}}) = (\pi_1(\boldsymbol{\mathcal{Z}}), \ldots, \pi_l(\boldsymbol{\mathcal{Z}}))^T$, and $\boldsymbol{\pi}_2(\boldsymbol{\mathcal{Z}}) = E(\mathbf{X}_2|\boldsymbol{\mathcal{Z}}) = (\pi_{l+1}(\boldsymbol{\mathcal{Z}}), \ldots, \pi_{l+p}(\boldsymbol{\mathcal{Z}}))^T$. To simplify notation, define

$$\boldsymbol{\Pi}_1(\boldsymbol{\mathcal{Z}}) = \begin{pmatrix} \mathbf{Z}_{12} \\ \boldsymbol{\pi}_1(\boldsymbol{\mathcal{Z}}) \end{pmatrix}, \quad \boldsymbol{\Pi}_2(\boldsymbol{\mathcal{Z}}) = \begin{pmatrix} \mathbf{Z}_{13} \\ \boldsymbol{\pi}_2(\boldsymbol{\mathcal{Z}}) \end{pmatrix}, \quad \boldsymbol{\Theta}_g(\mathbf{z}_{11}) = \begin{pmatrix} \mathbf{g}_1(\mathbf{z}_{11}) \\ \mathbf{g}_2(\mathbf{z}_{11}) \end{pmatrix}, \quad \boldsymbol{\Theta}_p = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix},$$

and $\boldsymbol{\Theta}(\mathbf{z}_{11}) = \begin{pmatrix} \boldsymbol{\Theta}_g(\mathbf{z}_{11}) \\ \boldsymbol{\Theta}_p \end{pmatrix}$, where $\boldsymbol{\Theta}_g(\mathbf{z}_{11})$ is a $(d+l)$-dimension vector of functional coefficients and $\boldsymbol{\Theta}_p$ is a $(m+p)$-dimension vector of constant coefficients, then, (2) becomes

$$E(Y|\boldsymbol{\mathcal{Z}}) = \boldsymbol{\Pi}_1^T \, \boldsymbol{\Theta}_g(\mathbf{z}_{11}) + \boldsymbol{\Pi}_2^T \, \boldsymbol{\Theta}_p = \boldsymbol{\Pi}_*^T \, \boldsymbol{\Theta}(\mathbf{Z}_{11}) \tag{3}$$

with $\boldsymbol{\Pi}_* = \begin{pmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2 \end{pmatrix}$.

Now we turn to the issue of identification. By the uniqueness of conditional expectations, $\{\boldsymbol{\pi}_j(\boldsymbol{\mathcal{Z}})\}$ in (2) are identified up to an additive constant. Thus, they are treated as identified components in the following discussion. It is easy to see from (3) that $\boldsymbol{\Theta}(\mathbf{Z}_{11}) = \boldsymbol{\Omega}_0(\mathbf{z}_{11})^{-1} \, E[\boldsymbol{\Pi}_* \, Y \,|\, \mathbf{Z}_{11} = \mathbf{z}_{11}]$, where $\boldsymbol{\Omega}_0(\mathbf{z}_{11}) = E\left\{\boldsymbol{\Pi}_* \, \boldsymbol{\Pi}_*^T \,|\, \mathbf{Z}_{11} = \mathbf{z}_{11}\right\}$. This implies that parameters $\boldsymbol{\Theta}_p$ are identified and coefficient functions $\boldsymbol{\Theta}_g\,(\mathbf{z}_{11})$ are identified up to an additive constant if $\boldsymbol{\Omega}_0(\mathbf{z}_{11})$ is positive definite for any $\mathbf{z}_{11}$.

From equation (3), $\boldsymbol{\Theta}$ can be estimated by a local linear regression of Y on $\boldsymbol{\mathcal{Z}}$ if $\boldsymbol{\Pi}_*$ would be known. However, $\{\boldsymbol{\pi}_j(\boldsymbol{\mathcal{Z}}),\ j = 1, 2\}$ are unknown functions in practice. Therefore, this leads to estimating $\boldsymbol{\pi}_j$ first by a regression of $\mathbf{X}_j$ on $\boldsymbol{\mathcal{Z}}$. The next step is to estimate $\boldsymbol{\Theta}$ by a regression of Y on $\boldsymbol{\mathcal{Z}}$ and the estimated values $\widehat{\boldsymbol{\pi}}_j$ obtained from the first step. Note that while $\boldsymbol{\Theta}_p$ is a global parameter, the estimation of $\boldsymbol{\Theta}_p$ only involves the local data points in a neighborhood of $\mathbf{Z}_{11}$. An efficient estimation of the constant coefficients requires using all data points. An easy way here is to use the average method to obtain an efficient estimator for $\boldsymbol{\Theta}_p$ which is shown to be $\sqrt{n}$-consistent although other alternative approaches might be applicable.

To estimate the functional coefficients $\boldsymbol{\Theta}_g(\cdot)$, let $\widehat{\boldsymbol{\beta}}_{1,n}$ and $\widehat{\boldsymbol{\beta}}_{2,n}$ denote any $\sqrt{n}$-consistent estimates of parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in (2). Define the estimated part residual as $\widehat{Y}^* = Y - \widehat{\boldsymbol{\beta}}_{1,n}^T \mathbf{Z}_{13} - \widehat{\boldsymbol{\beta}}_{2,n}^T \mathbf{X}_2$. Then, model (1) is approximated by

$$\widehat{Y}^* \approx \mathbf{g}_1(\mathbf{Z}_{11})^T \mathbf{Z}_{12} + \mathbf{g}_2(\mathbf{Z}_{11})^T \mathbf{X}_1 + \varepsilon,$$

which becomes a nonparametric functional coefficient IV model studied by CDXW (2006). Now, the nonparametric estimate of the nonparametric part $\mathbf{\Theta}_g(\cdot)$ can be obtained by applying the two-step nonparametric method proposed in CDXW (2006). It is shown that this nonparametric estimator is "oracle" in the sense that the asymptotic properties of this nonparametric estimator are not affected by knowing $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ or not. All details about estimation procedures and their properties are presented in Section 3.

3 Estimation Procedures

With the observed data $\{(Y_i, \mathbf{X}_i, \boldsymbol{\mathcal{Z}}_i)\}$, we propose the following procedures to obtain an efficient estimator for $\mathbf{\Theta}_p$ and to estimate the functional coefficients $\mathbf{\Theta}_g(\cdot)$ nonparametrically. In what follows, we apply a local linear fitting to estimate functionals although other smoothing methods such as the Nadaraya-Watson kernel method and spline methods are applicable. A local linear fitting is favored due to its attractive properties, such as high statistical efficiency in an asymptotic minimax sense, design adaptation, and automatic boundary effect corrections. The detailed description of this approach can be found in the book by Fan and Gijbels (1996) and its basic idea is illustrated next. Note that although a general local polynomial technique is applicable here, the local linear fitting might be enough for many applications according to Fan and Gijbels (1996) and the theory developed for the local linear estimator holds for the local polynomial estimator with some minor modifications. Therefore, for simplicity, the main focus here is only on the local linear estimation. To apply the local linear estimation procedure, it assumes commonly that all coefficient functions $\mathbf{g}_1(\cdot)$ and $\mathbf{g}_2(\cdot)$, and the regression functions $\boldsymbol{\pi}_1(\cdot)$ and $\boldsymbol{\pi}_2(\cdot)$ have a continuous second derivative. This assumption is made throughout the paper.

3.1 Efficient Estimation of Constant Coefficients

The first stage is to obtain the fitted value $\widehat{\boldsymbol{\pi}}_j$, $j = 1, 2$ by a nonparametric regression of $\mathbf{X}$ on $\boldsymbol{\mathcal{Z}}$. To this end, a combination of the local linear fitting and jack-knife technique is used. The local linear estimate of $\pi_j(\boldsymbol{\mathcal{Z}}_i)$ is denoted by $\widehat{\pi}_{j,-i}(\boldsymbol{\mathcal{Z}}_i)$ which is $\widehat{a}_j$, the minimizer of the following locally weighted least squares

$$\sum_{k \neq i} \{X_{kj} - a_j - \mathbf{b}_j^T(\boldsymbol{\mathcal{Z}}_k - \boldsymbol{\mathcal{Z}}_i)\}^2\, K_{h_1}(\boldsymbol{\mathcal{Z}}_k - \boldsymbol{\mathcal{Z}}_i),\; 1 \leq j \leq l + p,$$

where $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1^{d_1}$, $K(\cdot)$ is a kernel function on $\Re^{d_1}$, and $h_1 = h_{1n} > 0$ is the bandwidth at the first step, satisfying $h_1 \to 0$ and $n\, h_1^{d_1} \to \infty$ as $n \to \infty$. One advantage of using jack-knife technique is that the fitted value $\widehat{\pi}_{j,-i}(\boldsymbol{\mathcal{Z}}_i)$ is independent of $\mathbf{X}_i$. This can simplify the theoretical proofs although the asymptotic results are valid without using it.

At the second stage, we derive the estimation for constant coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. Since they are constant parameters, for simplicity, we apply the local constant (Nadaraya-Watson) estimation at this stage. By the continuity of $\{g_j(\mathbf{Z}_{11})\}$, for $\mathbf{Z}_{11}$ in a neighborhood of $\mathbf{z}_{11}$, $g_j(\mathbf{Z}_{11})$ is approximated by $g_j(\mathbf{z}_{11})$; that is, $g_j(\mathbf{Z}_{11}) \approx g_j(\mathbf{z}_{11})$. Define, for any $1 \le i \le n$, $\mathbf{\Pi}_{i1} = \mathbf{\Pi}_1(\boldsymbol{\mathcal{Z}}_i)$, $\mathbf{\Pi}_{i2} = \mathbf{\Pi}_2(\boldsymbol{\mathcal{Z}}_i)$, $\mathbf{\Pi}_{*,i} = \begin{pmatrix} \mathbf{\Pi}_{i1} \\ \mathbf{\Pi}_{i2} \end{pmatrix}$, $\bar{\mathbf{\Pi}} = (\mathbf{\Pi}_{*,1}, \ldots, \mathbf{\Pi}_{*,n})^T$, and $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, then, $E(\mathbf{Y}|\boldsymbol{\mathcal{Z}}_1, \cdots, \boldsymbol{\mathcal{Z}}_n) \approx \bar{\mathbf{\Pi}}\, \mathbf{\Theta}(\mathbf{z}_{11})$. Since $\pi_j(\boldsymbol{\mathcal{Z}}_i)$ in $\bar{\mathbf{\Pi}}$ is unobservable, we replace it by its fitted value $\widehat{\pi}_{j,-i}(\boldsymbol{\mathcal{Z}}_i)$ obtained from the previous step. A local constant estimator of $\widehat{\mathbf{\Theta}}$ at each sample point $\boldsymbol{\mathcal{Z}}_i$ can be obtained by minimizing the following locally weighted least squares

$$\sum_{s \neq i}^{n} \left[Y_s - \widehat{\mathbf{\Pi}}_{1,-s}\, \mathbf{\Theta}_g(\mathbf{Z}_{i11}) - \widehat{\mathbf{\Pi}}_{2,-s}\, \mathbf{\Theta}_p(\mathbf{Z}_{i11}) \right]^2 L_{h_2}(\mathbf{Z}_{s11} - \mathbf{Z}_{i11}), \tag{4}$$

where $\widehat{\mathbf{\Pi}}_{1,-s}$ is similar to $\mathbf{\Pi}_{s1}$ with $\pi_j(\boldsymbol{\mathcal{Z}}_s)$ replaced by $\widehat{\pi}_{j,-s}(\boldsymbol{\mathcal{Z}}_s)$, $L_{h_2}(\cdot) = L(\cdot/h_2)/h_2^k$, $L(\cdot)$ is a kernel function on $\Re^k$, and $h_2 = h_{2n} > 0$ is the bandwidth at the second step, satisfying $h_2 \to 0$ and $n\, h_2^k \to \infty$ as $n \to \infty$.

Let $\widehat{\bar{\mathbf{\Pi}}}$ be $\bar{\mathbf{\Pi}}$ with $\pi_j(\boldsymbol{\mathcal{Z}}_i)$ replaced by $\widehat{\pi}_{j,-i}(\boldsymbol{\mathcal{Z}}_i)$, and define $\mathbf{W}(\mathbf{z}_{11}) = \mathrm{diag}\{L_{h_2}(\mathbf{Z}_{111} - \mathbf{z}_{11}), \cdots, L_{h_2}(\mathbf{Z}_{n11} - \mathbf{z}_{11})\}$ and $\mathbf{W}_i = \mathbf{W}(\mathbf{Z}_{i11})$. It is not hard to verify that the minimizers of (4) are given by

$$\widehat{\mathbf{\Theta}}(\mathbf{Z}_{i11}) = \left(\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i\, \widehat{\bar{\mathbf{\Pi}}} \right)^{-1} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i\, \mathbf{Y}. \tag{5}$$

Hence the estimator of the constant coefficient at the sample point $\mathbf{Z}_{i11}$ is given by

$$\widehat{\mathbf{\Theta}}_p(\mathbf{Z}_{i11}) = \mathbf{e}^T\, \widehat{\mathbf{\Theta}}(\mathbf{Z}_{i11})$$

where $\mathbf{e} = (\,\mathbf{0} \quad \mathbf{I}_{(m+p)}\,)^T$ is a $(d+l+m+p) \times (m+p)$ matrix. Now the problem arises for $\widehat{\mathbf{\Theta}}_p(\mathbf{Z}_{i11})$ is that this approach only uses data points in a local neighborhood of $\mathbf{Z}_{i11}$ so that it might not be efficient since the constant coefficients $\mathbf{\Theta}_p$ are indeed global parameters. To overcome this problem, finally, the third step is to use the average method to estimate $\mathbf{\Theta}_p$ as follows

$$\widehat{\mathbf{\Theta}}_p = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{\Theta}}_p(\mathbf{Z}_{i11}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}^T \left(\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i\, \widehat{\bar{\mathbf{\Pi}}} \right)^{-1} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i\, \mathbf{Y} \tag{6}$$

although other alternative methods might be applicable. Note that the programming involved in the above three stage estimation is relatively simple and can be modified with few efforts from the existing programs for a linear IV model with a weight.

3.2 Nonparametric Estimation of Functional Coefficients

Next, we present an estimator of the functional coefficients at any point $\mathbf{z}_{11}$ if any $\sqrt{n}$ consistent estimator $\widehat{\boldsymbol{\Theta}}_p$ is given. It follows from (2) that

$$E(Y - \boldsymbol{\beta}_1^T \mathbf{Z}_{13} - \boldsymbol{\beta}_2^T \mathbf{X}_2 | \boldsymbol{\mathcal{Z}}) = \mathbf{g}_1(\mathbf{Z}_{11})^T \mathbf{Z}_{12} + \mathbf{g}_2(\mathbf{Z}_{11})^T \boldsymbol{\pi}_1(\boldsymbol{\mathcal{Z}}). \tag{7}$$

Then, the right hand side of equation (7) has the functional coefficient form as in CDXW (2006) so that the estimation method proposed by CDXW (2006) can be applied to (7) with some minor changes. Since $\{g_j(\cdot)\}$ have a continuous second derivative at any point $\mathbf{z}_{11}$, for $\mathbf{Z}_{11}$ in a neighborhood of $\mathbf{z}_{11}$, an application of the Taylor expansion gives $g_j(\mathbf{Z}_{11}) \approx g_j(\mathbf{z}_{11}) + (\mathbf{Z}_{11} - \mathbf{z}_{11})^T g_j'(\mathbf{z}_{11})$, where $g_j'(\cdot)$ is the first partial derivative of $g_j(\cdot)$. Let $\otimes$ be the Kronecker product and define

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}_{11}^T & \boldsymbol{\Pi}_{11}^T \otimes (\mathbf{Z}_{111} - \mathbf{z}_{11})^T \\ \vdots & \vdots \\ \boldsymbol{\Pi}_{n1}^T & \boldsymbol{\Pi}_{n1}^T \otimes (\mathbf{Z}_{n11} - \mathbf{z}_{11})^T \end{pmatrix}.$$

Then, (7) can be approximated by $E(\widetilde{\mathbf{Y}} | \boldsymbol{\mathcal{Z}}_1, \cdots, \boldsymbol{\mathcal{Z}}_n) \approx \boldsymbol{\Pi}\, \widetilde{\boldsymbol{\Theta}}$, where the partial residuals $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \cdots, \widetilde{Y}_n)^T$, $\widetilde{Y}_i = Y_i - \boldsymbol{\beta}_1^T \mathbf{Z}_{i13} - \boldsymbol{\beta}_2^T \mathbf{X}_{i2}$, and $\widetilde{\boldsymbol{\Theta}} = \widetilde{\boldsymbol{\Theta}}(\mathbf{z}_{11}) = \begin{pmatrix} \boldsymbol{\Theta}_g(\mathbf{z}_{11}) \\ \boldsymbol{\Theta}_g'(\mathbf{z}_{11}) \end{pmatrix}$ with $\boldsymbol{\Theta}_g'(\cdot)$ being the first partial derivative of $\boldsymbol{\Theta}_g(\cdot)$. Since $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\pi}_1$ are unknown, they are replaced by their fitted values obtained from previous steps. The local linear estimation of $\widetilde{\boldsymbol{\Theta}}$ is the minimizer of the following locally weighted least squares

$$\begin{aligned} &\sum_{i=1}^{n} \left[\widehat{\widetilde{Y}}_i - \sum_{j=1}^{d} \left\{ g_j(\mathbf{z}_{11}) + (\mathbf{Z}_{i11} - \mathbf{z}_{11})^T g_j'(\mathbf{z}_{11}) \right\} Z_{j,\,i12} \right. \\ &\left. - \sum_{j=d+1}^{d+l} \left\{ g_j(\mathbf{z}_{11}) + (\mathbf{Z}_{i11} - \mathbf{z}_{11})^T g_j'(\mathbf{z}_{11}) \right\} \widehat{\pi}_{j,\,-i}(\boldsymbol{\mathcal{Z}}_i) \right]^2 M_{h_3}(\mathbf{Z}_{i11} - \mathbf{z}_{11}), \end{aligned} \tag{8}$$

where $\widehat{\widetilde{Y}}_i = Y_i - \widehat{\boldsymbol{\beta}}_1^T \mathbf{Z}_{i13} - \widehat{\boldsymbol{\beta}}_2^T \mathbf{X}_{i2}$ is the estimated partial residual, $M_{h_3}(\cdot) = M(\cdot/h_3)/h_3$, $M(\cdot)$ is a kernel function on $\Re^k$, and $h_3 = h_{3n} > 0$ is the bandwidth satisfying $h_3 \to 0$ and $n\,h_3^k \to \infty$ as $n \to \infty$.

Let $\widehat{\mathbf{\Pi}}$ be defined as same as $\mathbf{\Pi}$ with $\pi_j(\boldsymbol{\mathcal{Z}}_i)$ replaced by $\widehat{\pi}_{j,-i}(\boldsymbol{\mathcal{Z}}_i)$, then it is not hard to show that the minimizers of (8) are given by

$$\widehat{\widehat{\mathbf{\Theta}}} = \left(\widehat{\mathbf{\Pi}}^T \widetilde{\mathbf{W}}\, \widehat{\mathbf{\Pi}}\right)^{-1} \widehat{\mathbf{\Pi}}^T \widetilde{\mathbf{W}}\, \widehat{\widehat{\mathbf{Y}}},$$

where $\widetilde{\mathbf{W}} = \mathrm{diag}\{M_{h_3}(\mathbf{Z}_{111} - \mathbf{z}_{11}), \cdots, M_{h_3}(\mathbf{Z}_{n11} - \mathbf{z}_{11})\}$ and $\widehat{\widehat{\mathbf{Y}}} = (\widehat{\widehat{Y}}_1, \ldots, \widehat{\widehat{Y}}_n)^T$. In particular, the local linear estimates of the coefficient functions are given by

$$\widehat{g}_j(\mathbf{z}_{11}) = \mathbf{e}_j^T\, \widehat{\widehat{\mathbf{\Theta}}}, \quad 1 \le j \le d + l,$$

where $\mathbf{e}_j$ is a $(d + l)(k + 1)$ vector with the j-th element being 1 and the remaining elements zero.

4 Distribution Theory

4.1 Assumptions and Notation

In this section, we derive the consistency and asymptotic normality of the constant and functional coefficients estimators proposed above. First, introduce some notations. For simplicity, arguments are sometimes dropped. Let $f_{11}(\cdot)$ be the probability density function of $\mathbf{Z}_{11}$, $\boldsymbol{\xi} = \mathbf{X} - E(\mathbf{X}\,|\,\boldsymbol{\mathcal{Z}})$, and $\eta = Y - E(Y\,|\,\boldsymbol{\mathcal{Z}})$. Also, set $\sigma_\eta^2(\boldsymbol{\mathcal{Z}}) = E(\eta^2|\boldsymbol{\mathcal{Z}})$, $\mathbf{\Sigma}_\xi(\boldsymbol{\mathcal{Z}}) = E(\boldsymbol{\xi}\boldsymbol{\xi}^T|\boldsymbol{\mathcal{Z}})$, and $\mathbf{\Sigma}_{\eta\xi}(\boldsymbol{\mathcal{Z}}) = E(\eta\,\boldsymbol{\xi}|\boldsymbol{\mathcal{Z}})$ To characterize the asymptotic variance of estimators $\widehat{\mathbf{\Theta}}_p$ of constant coefficients, define

$$\mathbf{\Sigma}_p = E\left\{\mathbf{e}^T\, \mathbf{\Omega}_0^{-1}(\mathbf{\Omega}_\eta + \mathbf{\Omega}_\xi - 2\,\mathbf{\Omega}_{\eta\xi})\,\mathbf{\Omega}_0^{-1}\mathbf{e}\right\}, \tag{9}$$

where $\mathbf{\Omega}_\pi(\boldsymbol{\mathcal{Z}}) \equiv \mathbf{\Pi}_*(\boldsymbol{\mathcal{Z}})\ \mathbf{\Pi}_*(\boldsymbol{\mathcal{Z}})^T$, $\mathbf{\Omega}_\eta = \mathbf{\Omega}_\pi\, \sigma_\eta^2(\boldsymbol{\mathcal{Z}})$, $\mathbf{\Theta}_* = \begin{pmatrix} \mathbf{g}_2 \\ \boldsymbol{\beta}_2 \end{pmatrix}$, $\mathbf{\Omega}_\xi = \mathbf{\Omega}_\pi\, \mathbf{\Theta}_*^T\, \mathbf{\Sigma}_\xi\, \mathbf{\Theta}_*$, and $\mathbf{\Omega}_{\eta\,\xi} = \mathbf{\Omega}_\pi\, \mathbf{\Sigma}_{\eta\xi}^T\, \mathbf{\Theta}_*$. Further, to present the asymptotic variance of estimator $\widehat{\mathbf{\Theta}}_g$ of functional coefficients, define $\boldsymbol{\mu}_2(M) = \int \mathbf{u}\,\mathbf{u}^T\, M(\mathbf{u})\, d\mathbf{u}$ and $\nu_0(M) = \int M^2(\mathbf{u})\, d\mathbf{u}$. Let $\mathbf{\Sigma}_{\xi,11}$ be the first block of $\mathbf{\Sigma}_\xi$, $\mathbf{\Sigma}_{\eta\xi,1}$ denote the first part of $\mathbf{\Sigma}_{\eta\xi}$, and $\mathbf{\Omega}_{\pi,11}(\boldsymbol{\mathcal{Z}})$ present the first block of $\mathbf{\Omega}_\pi(\boldsymbol{\mathcal{Z}})$. Further, denote $\widetilde{\mathbf{\Omega}}_0(\mathbf{z}_{11}) = E\left\{\mathbf{\Omega}_{\pi,11}(\boldsymbol{\mathcal{Z}})\,|\mathbf{Z}_{11} = \mathbf{z}_{11}\right\}$,

$$\widetilde{\mathbf{\Omega}}_\eta = E\left\{\mathbf{\Omega}_{\pi,11}(\boldsymbol{\mathcal{Z}})\,\sigma_\eta^2(\boldsymbol{\mathcal{Z}})|\mathbf{Z}_{11} = \mathbf{z}_{11}\right\}\,\nu_0(M),$$

$$\widetilde{\mathbf{\Omega}}_\xi = E\left\{\mathbf{\Omega}_{\pi,11}(\boldsymbol{\mathcal{Z}})\mathbf{\Theta}_g^T(\mathbf{z}_{11})\mathbf{\Sigma}_{\xi,11}\mathbf{\Theta}_g(\mathbf{z}_{11})|\mathbf{Z}_{11} = \mathbf{z}_{11}\right\}\,\nu_0(M),$$

and

$$\widetilde{\mathbf{\Omega}}_{\eta\,\xi} = E\{\mathbf{\Omega}_{\pi,11}(\boldsymbol{\mathcal{Z}})\mathbf{\Sigma}_{\eta\xi,1}\mathbf{\Theta}_g(\mathbf{z}_{11})|\mathbf{Z}_{11} = \mathbf{z}_{11}\}\,\nu_0(M).$$

Finally, set $\mathbf{\Sigma}_g(\mathbf{z}_{11}) = f_{11}^{-1}(\mathbf{z}_{11})\,\widetilde{\mathbf{\Omega}}_0^{-1}(\mathbf{z}_{11})\,\widetilde{\mathbf{\Omega}}_1(\mathbf{z}_{11})\,\widetilde{\mathbf{\Omega}}_0^{-1}(\mathbf{z}_{11})$, where $\widetilde{\mathbf{\Omega}}_1(\mathbf{z}_{11}) = \widetilde{\mathbf{\Omega}}_\eta + \widetilde{\mathbf{\Omega}}_\xi - 2\,\widetilde{\mathbf{\Omega}}_{\eta\xi}$. The following conditions are collected together for our asymptotic theory.

Assumptions:

1. The kernels $K(\cdot)$, $L(\cdot)$, and $M(\cdot)$ are symmetric and bounded second order kernel functions with bounded support. Further, $K(\cdot)$ and $L(\cdot)$ satisfy a Lipschitz condition of degree one.

2. The density function $f(\cdot)$ of $\boldsymbol{\mathcal{Z}}$ is bounded and uniformly continuous and there exists a compact set D such that $\inf_{\mathbf{z}\in D} f_{11}(\mathbf{z}) > 0$.

3. Functions $\{g_j''(\cdot)\}$, $\sigma_\eta^2(\cdot)$, $\mathbf{\Sigma}_\xi(\cdot)$, and $\mathbf{\Sigma}_{\eta\xi}(\cdot)$ are continuous, where $g_j''(\cdot)$ is the second partial derivative of $g_j(\cdot)$. Further, $\{\pi_j''(\cdot)\}$ are bounded and uniformly continuous and satisfy the Lipschitz condition where $\pi_j''(\cdot)$ is the second partial derivative of $\pi_j(\cdot)$.

4. Observations $\{(\mathbf{X}_i, Y_i, \boldsymbol{\mathcal{Z}}_i)\}_{i=1}^n$ are independent and identically distributed, the fourth moment of $\boldsymbol{\xi}$ and η exists.

5. $\mathbf{\Omega}_0(\mathbf{z}_1)$ is positive definite.

6. $E\,|\mathbf{X}_j|^\gamma < \infty$ for some $\gamma > 2$.

7. $n\,h_1^{d_1\alpha_1}/\log h_1 \to \infty$ with $\alpha_1 > \gamma/(\gamma-2)$, $h_1 \to 0$, $n\,h_2^{k\,\alpha_2}/\log h_2 \to \infty$ with $\alpha_2 > \gamma/(\gamma-2)$, $h_2 \to 0$, $n\,h_3^k \to \infty$, and $h_3 \to 0$.

8. $h_1 = o(h_2)$ and $h_2 = o(h_3)$.

Remark 1: (*Discussion of Conditions*) Assumptions 1-8 are similar to those for varying coefficient IV models in CDXW (2006) and they are fairly standard. Since an additional step is needed to estimate the constant coefficient $\mathbf{\Theta}_p$, assumptions are needed on three kernel functions and three bandwidths, respectively. The Lipschitz condition on kernels $K(\cdot)$ and $L(\cdot)$ is a technical assumption to simplify the theoretical proofs but it might not be necessary. Under this condition, the uniform convergence is achieved at both the first and second steps. Note that Assumption 5 is also the sufficient condition for the model identification as in Section 2. Assumption 8 suggests that the first two steps should be under-smoothed given that Assumption 7 is satisfied so that the biases from the earlier stages are negligible at the later stages. Also note that similar conditions are imposed for the two-stage method for ordinary regression models for cross-sectional and time series data; see Cai (2002a, 2002b).

4.2 Asymptotic Properties

We now present the consistency and asymptotic normality for $\widehat{\mathbf{\Theta}}_p$, the estimator of constant coefficients.

Theorem 1 . *Under Assumptions 1-8, one has*

$$\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p - bias_p = o_p(h_2^2) + O_p(n^{-1/2}),$$

where

$$bias_p = \frac{h_2^2}{2}\mathbf{e}^T E\left\{\mathbf{\Omega}_0^{-1}\,\mathbf{\Omega}_\pi(\boldsymbol{\mathcal{Z}})\begin{pmatrix} \left(tr\left[\boldsymbol{\mu}_2(L)\left\{2\,g_j'(\mathbf{Z}_{11})\,f_{11}'(\mathbf{Z}_{11})^T/f_{11}(\mathbf{Z}_{11}) + g_j''(\mathbf{Z}_{11})\right\}\right]\right)_{(d+l)\times 1} \\ \mathbf{0}_{(m+p)\times 1}\end{pmatrix}\right\}.$$

Theorem 2 . *Under Assumptions 1-8, if $n\,h_2^4 = O(1)$, we have*

$$\sqrt{n}\,\left[\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p - bias_p\right] \;\to\; N(\mathbf{0}, \mathbf{\Sigma}_p).$$

Particularly, if $n\,h_2^4 = o(1)$, then,

$$\sqrt{n}\,\left[\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p\right] \;\to\; N(\mathbf{0}, \mathbf{\Sigma}_p).$$

Remark 2: As seen in Theorems 1 and 2, the bias of our final estimators for constant coefficients is of the order of h_2^2, while the variance is of the order of $1/n$. When $n\,h_2^4 \to 0$, the estimators achieve $1/\sqrt{n}$, the optimal rate for parametric models. Since the local constant estimation is used in the first step, the bias comes from the first and second derivatives of the functional coefficients $g_j(\cdot)$ due to the approximation errors. The approximation error of $\pi_j(\cdot)$ from the first step is negligible because of Assumption 8. Due to the average in (6), the variance of $\widehat{\mathbf{\Theta}}_p$ reduces to the order of $1/n$ but not depend on h_2. Asymptotically, the effect of the first step bandwidth on the variance does not carry over to the second step. Note that, like $\mathbf{\Sigma}_g$ (the asymptotic variance of $\widehat{\mathbf{\Theta}}_g$; see Theorem 4 later), $\mathbf{\Sigma}_p$ (the asymptotic variance of $\widehat{\mathbf{\Theta}}_p$) also includes three components: the variance of reduced form measurement error, the variance of the structural equation measurement error, and the covariance between them.

Next, it is shown that $\widehat{\mathbf{\Theta}}_g$, the estimators for the functional coefficients share the same asymptotic properties as the estimators in the nonparametric varying coefficient IV models discussed in CDXW (2006). In other words, the estimators perform asymptotically as well as if the constant coefficients were known. This is referred as an efficient or oracle estimator.

Theorem 3 . *Under Assumptions 1-8, then,*

$$\widehat{\mathbf{\Theta}}_g - \mathbf{\Theta}_g - bias_g(\mathbf{z}_{11}) = o_p\left(h_3^2\right) + O_p\left\{(n\, h_3^k)^{-1/2}\right\},$$

where

$$bias_g(\mathbf{z}_{11}) = \frac{h_3^2}{2}\left(tr\left\{\boldsymbol{\mu}_2(M) g_j''(\mathbf{z}_{11})\right\}\right)_{(d+l)\times 1}.$$

Theorem 4 . *Under Assumptions 1-8, if* $nh_3^{k+4} = O(1)$, *then,*

$$\sqrt{nh_3^k}\left[\widehat{\mathbf{\Theta}}_g(\mathbf{z}_{11}) - \mathbf{\Theta}_g(\mathbf{z}_{11}) - bias_g(\mathbf{z}_{11})\right] \longrightarrow N(\mathbf{0},\ \mathbf{\Sigma}_g(\mathbf{z}_{11})).$$

Remark 3: Theorems 3 and 4 demonstrate that the asymptotic properties of our final estimators for functional coefficients in the partially varying coefficient IV model are the same as those for the estimators in the nonparametric varying coefficient IV model discussed in CDXW (2006). Because of Assumption 8, the biases from previous steps are negligible comparing to the final stage bias at the order of h_3^2. Similar to the aforementioned discussion, the effect of the earlier stage bandwidths on variance asymptotically does not carry over to the final stage. In fact, asymptotically, the only difference between $\widehat{\mathbf{\Theta}}_g$ and the estimator for the nonparametric varying coefficient IV model in CDXW (2006) relies on the extra term which is shown to be negligible due to the faster convergence rate of $\widehat{\mathbf{\Theta}}_p$ (see Appendix for details). Therefore, the final estimator $\widehat{\mathbf{\Theta}}_g$ is optimal in the sense that it performs as if those constant coefficients were known.

Remark 4: (*Bandwidth Selection*) Intuitively, constant coefficients are global parameters so that a large bandwidth is preferred. On the other hand, a small bandwidth is desirable for functional coefficients to reduce bias. To reconcile this contradiction, we first employ a small bandwidth to reduce the bias and then take the average of all local estimates. In such a way, the variance of the constant coefficient estimators is stabilized although a small bias is preserved. Besides, by observing Theorems 1 and 2, we notice that the variance term does not involve h_2. Therefore, the first and second step bandwidths should be chosen as small as possible as long as they satisfy Assumptions 7 and 8. An ad hoc bandwidth selection method discussed in CDXW (2006) for functional coefficient IV model is applicable here for the partially varying coefficient IV model; see CDXW (2006) for details. Finally, remark that there appears to be no any result available in the literature for a data-driven bandwidth selection with optimal properties. It is still an open question for future work, but it is beyond the scope of the present paper to give a more precise result. Nevertheless, the procedure suggested above is a useful one for practitioners and found to be practicable in our own empirical examples in Section 5.

4.3 Covariance Matrix Estimation

Large sample confidence intervals are useful for inference. Next, we consider a consistent estimation of the asymptotic covariance matrix for the construction of confidence intervals.

For $\widehat{\boldsymbol{\Theta}}_p$ (the estimator of constant coefficients), a universal confidence interval is desirable. In view of (9), a direct estimate of $\boldsymbol{\Sigma}_p$ can be constructed. By Lemma 2 in Section 6, it follows that at each data point $\mathbf{Z}_{s11}$,

$$\frac{1}{n}\widehat{\widehat{\boldsymbol{\Pi}}}^T\mathbf{W}_s\widehat{\widehat{\boldsymbol{\Pi}}} = \frac{1}{n}\sum_i \widehat{\boldsymbol{\Omega}}_{\pi,-i}L_{h_2}(\mathbf{Z}_{i11}-\mathbf{Z}_{s11}) = f_{11}(\mathbf{Z}_{s11})\,\boldsymbol{\Omega}_0(\mathbf{Z}_{s11}) + o_p\,(1),$$

with

$$\widehat{\boldsymbol{\Omega}}_{\pi,-i} = \begin{pmatrix} \widehat{\boldsymbol{\Pi}}_{1,-i}\widehat{\boldsymbol{\Pi}}_{1,-i}^T & \widehat{\boldsymbol{\Pi}}_{1,-i}\widehat{\boldsymbol{\Pi}}_{2,-i}^T \\ \widehat{\boldsymbol{\Pi}}_{2,-i}\widehat{\boldsymbol{\Pi}}_{1,-i}^T & \widehat{\boldsymbol{\Pi}}_{2,-i}\widehat{\boldsymbol{\Pi}}_{2,-i}^T \end{pmatrix}.$$

Also, observe that

$$\widehat{f}_{11}(\mathbf{Z}_{s11}) = \frac{1}{n}\sum_i L_{h_2}\left(\mathbf{Z}_{i11}-\mathbf{Z}_{s11}\right)$$

is a consistent estimate of $f_{11}(\mathbf{Z}_{s11})$. Thus, $\boldsymbol{\Omega}_0(\mathbf{Z}_{s11})$ can be estimated consistently by

$$\widehat{\boldsymbol{\Omega}}_0(\mathbf{Z}_{s11}) = \frac{1}{n}\sum_i \widehat{\boldsymbol{\Omega}}_{\pi,-i}\,L_{h_2}(\mathbf{Z}_{i11}-\mathbf{Z}_{s11})\,/\,\widehat{f}_{11}(\mathbf{Z}_{s11}).$$

Similarly, a consistent estimate of other components in (9) can be obtained by

$$\widehat{\boldsymbol{\Omega}}_\eta(\mathbf{Z}_{s11}) = \frac{1}{n}\sum_i \widehat{\boldsymbol{\Omega}}_{\pi,-i}\,\widehat{\eta}_i^2 L_{h_2}(\mathbf{Z}_{i11}-\mathbf{Z}_{s11})\,/\,\widehat{f}_{11}(\mathbf{Z}_{s11}),$$

$$\widehat{\boldsymbol{\Omega}}_\xi(\mathbf{Z}_{s11}) = \frac{1}{n}\sum_i \widehat{\boldsymbol{\Omega}}_{\pi,-i}\,\widehat{\boldsymbol{\Theta}}_*^T\,\widehat{\boldsymbol{\xi}}_i\widehat{\boldsymbol{\xi}}_i^T\,\widehat{\boldsymbol{\Theta}}_*\,L_{h_2}(\mathbf{Z}_{i11}-\mathbf{Z}_{s11})\,/\,\widehat{f}_{11}(\mathbf{Z}_{s11}),$$

and

$$\widehat{\mathbf{\Omega}}_{\eta\,\xi}(\mathbf{Z}_{s11}) = \frac{1}{n}\,\sum_i \widehat{\mathbf{\Omega}}_{\pi,\,-i}\,\widehat{\eta}_i\,\widehat{\boldsymbol{\xi}}_i^T\,\widehat{\mathbf{\Theta}}_*\,L_{h_2}(\mathbf{Z}_{i11} - \mathbf{Z}_{s11})\,/\,\widehat{f}_{11}(\mathbf{Z}_{s11}),$$

with $\widehat{\eta}_i = Y_i - \widehat{\mathbf{\Pi}}_{1,\,-i}\widehat{\mathbf{\Theta}}_g(\mathbf{Z}_{i11}) - \widehat{\mathbf{\Pi}}_{2,\,-i}\widehat{\mathbf{\Theta}}_p$ and $\widehat{\boldsymbol{\xi}}_i = \mathbf{X}_i - \widehat{\boldsymbol{\pi}}_{-i}$. Finally, $\mathbf{\Sigma}_p$ is estimated by

$$\widehat{\mathbf{\Sigma}}_p = \frac{1}{n}\sum_s \mathbf{e}^T\,\left\{\widehat{\mathbf{\Omega}}_0(\mathbf{Z}_{s11})\right\}^{-1}\left\{\widehat{\mathbf{\Omega}}_\eta(\mathbf{Z}_{s11}) + \widehat{\mathbf{\Omega}}_\xi(\mathbf{Z}_{s11}) - 2\,\widehat{\mathbf{\Omega}}_{\eta\,\xi}(\mathbf{Z}_{s11})\right\}\left\{\widehat{\mathbf{\Omega}}_0(\mathbf{Z}_{s11})\right\}^{-1}\,\mathbf{e}.$$

The asymptotic covariance matrix of $\widehat{\mathbf{\Theta}}_g$ can be estimated by using the same expressions in CDXW (2006) with some minor changes of the error terms accordingly. Both covariance matrix estimators is examined by empirical examples in Section 5. Indeed, it is easy to show that the aforementioned estimators are consistent.

5 Empirical Examples

To illustrate the estimation procedure and investigate the finite sample performance of the proposed estimators, we consider one simulated example and revisit the random sample of young Australian female workers introduced in Section 1. We use the Epanechnikov kernel $K(u) = 0.75\,(1 - u^2)\,I(|u| \le 1)$ and its product form as the multivariate kernel and choose the bandwidths based on the ad hoc approach as described in Remark 4 in Section 4.2. For the simulated example, the finite sample performances of the our estimator are evaluated in terms of the mean absolute deviation error (MADE) defined by

$$\mathcal{E}_j = \frac{1}{n_0}\,\sum_{k=1}^{n_0} |\widehat{g}_j(z_k) - g_j(z_k)|$$

for $g_j(\cdot)$, where z_k, $k = 1, \cdots, n_0$ are the regular grid points.

5.1 A Real Example

First, we apply the proposed model and its estimation procedures to the random sample of young Australian female workers from the 1985 wave of the Australian Longitudinal Survey data. It is the same data that was studied by CDXW (2006). Although the actually fitted model is the same, the main focus of that paper was the functional coefficient part. Discussion of the constant parameter estimators, including their efficiency and asymptotic properties, was out of that paper's scope. Following we will pay special attention to those topics.

It is well known in the labor economics literature that, due to unobservable heterogeneity in schooling choices, education is an endogenous variable in the wage model, see, e.g., the review paper by Card (2001). Economists have also noted that the positive marginal returns to education vary with the level of schooling, see, e.g., Schulz (1997) for details. Further, if the work experience is also an attribute valued by employers, then, for any given level of education, wages should be increasing in experience. Card (2001) has suggested that if a wage model assumes the additive separability of education and experience, then returns to education could be understated if returns to experience is itself increasing in education.

The above features of the wage problem makes varying coefficient IV model a natural fit to the data. However, by observing the data, we also found four discrete exogenous variables have significant impact on the wage values. To incorporate those variables into the model, we consider the partially varying coefficient IV model proposed in this paper

$$Y = \mathbf{Z}_3^T \boldsymbol{\beta} + g_1(Z_1) + g_2(Z_1)\, X + \varepsilon, \tag{10}$$

and $E(X|Z_1,\, Z_2) = \pi_1(Z_1,\, Z_2)$, where Y is the natural logarithm of the hourly wage, $\mathbf{Z}_3$ includes binary indicators for marital status, government employed, union status, and Australian-born, Z_1 is a measure of work experience measured in years, X is the measure of (endogenous) education ("Schooling"), and Z_2 is an index of labor market attitudes that used as an instrumental variable. $g_1(\cdot)$ and $g_2(\cdot)$ are unknown coefficient functions. Here we follow Das, Newey and Vella (2003) to use the labor market attitude index as the instrumental variable. In that paper, they conducted a statistical test and rejected the possibility of endogeneity of the attitude index. The sample consists of 1996 observations.

The estimation results are summarized in Table 1 for the constant coefficients, and Figure

Table 1: Estimates of the Constant Coefficients in Structural Equation (10)

Variables	Marital Status	Government Employed	Union Status	Australian Born
Estimated Values	0.1523	0.6129	0.0734	-0.0825
Standard Errors	0.0300	0.0199	0.0190	0.0309

1 for the functional coefficient of schooling. The bandwidths used for estimating the reduced form are $h_{11} = 1.5$ for Z_1 and $h_{12} = 0.6$ for Z_2 in the first stage, $h_2 = 4$ in the second stage for constant coefficients, and $h_3 = 5$ in the third stage for functional coefficients.

Table 1 presents the estimates of constant coefficients $\boldsymbol{\beta}$ with associated standard errors. It can be seen that the estimates of standard errors are in line with the asymptotic properties established in Section 4.3. Based on that, it is clear that all coefficient estimates are significantly different from 0 at a 5% significance level based on a simple t-test. The interpretation is that individuals who are married, government employed, in trade union, or not born in Australian tend to have a higher wage. Married females are more mature and potentially have more experience, which is one of the attributes valued by employers. The officials in government generally get a better compensation. Union workers earn a higher wage. Aliens' average salary is higher than the natives. Considering Australia is one of the biggest immigrant country in the world, many immigrants come here to pursuit a higher education and get better jobs after they graduate. We also noticed among all these features, marital status and government employment have a bigger impact on wages.

Figure 1 plots the three-step local linear estimate of the functional coefficient $g_2(\cdot)$ correcting for endogeneity (smooth solid line) with 95% pointwise confidence intervals (dotted lines) with the bias ignored, and the ordinary local linear estimate without correcting for endogeneity (dashed line). There are several notable points about this figure. First, while the profile without correcting for endogeneity is almost constant, the profile correcting for endogeneity is positive and nonlinear for all values of experience in our sample. As in our simulated example (see next), this figure illustrates both the practicability of our estimators and the importance of correcting for endogeneity. Secondly, notice that the derivative of $g_2(\cdot)$ changes over its range, being negative at both low and high levels of experience but positive in the middle range. This suggests that while the marginal returns to education are positive, these returns are themselves declining in experience for both low and high level workers. Our results also show that the partially varying coefficient model captures the unknown nonlinear effect of education on wages discussed in Card (2001) and illustrates the practicability of the standard error estimators leading to the plotted 95 percent confidence intervals. Finally, by comparing our result with the model without parametric part (see Xiong (2004)), although the shape of $\widehat{g}_2(\cdot)$ is similar, $\widehat{g}_2(\cdot)$ from (10) is flatter in the middle range yet steeper at the higher end of experience span because some of the variations are dragged out by the additional parametric part in (10).

In summary, a partially varying coefficient model provides a very simple yet powerful modeling framework for the wage-education relationship. In addition to the extremely practicable functional coefficient specification, it also allows to incorporate more information in a parametric form without affecting the convergence rate on either parametric or nonparametric part. Our results also illustrate the three stage estimation procedure and covariance matrix estimation for inference purpose.

5.2 A Simulated Example

Next, we use the simulated data to illustrate our model and its estimation procedures. The data are generated through the following model

$$Y = g_1(Z_1) + g_2(Z_1)\, Z_2 + g_3(Z_1)\, X_1 + \beta_1\, Z_3 + \beta_2\, X_2 + \varepsilon,$$

where the coefficient functions $g_j(\cdot)$, $1 \le j \le 3$, are given by

$$g_1(z) = 2\cos(z) + 0.8268, \quad g_2(z) = -(2 + 0.2\, z) \exp\left\{-(0.5\, z - 1.5)^2\right\},$$

and $g_3(z) = \sin(z)$. The constant coefficients are $\beta_1 = 1$ and $\beta_2 = -1$. The exogenous variable Z_1 and instrumental variable Z_4 are independently generated from uniform distribution (2, 8), and exogenous variables Z_2 and Z_3 are independently generated from normal distribution $N(0,\, 1)$. The endogenous variables X_1 and X_2 are generated by

$$X_1 = 3\sin(2\, Z_4) + \xi_1, \quad \text{and} \quad X_2 = 3\, \cos(2\, Z_4) + \xi_2.$$

The error distributions ε, ξ_1, and ξ_2 are generated jointly from a multivariate normal distribution as

$$\begin{pmatrix} \varepsilon \\ \xi_1 \\ \xi_2 \end{pmatrix} \sim \mathbf{N}\left(\mathbf{0},\, \begin{pmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0 \\ 0.7 & 0 & 1 \end{pmatrix}\right).$$

Clearly, ε is independent of Z_1 and Z_2 so that $E(\varepsilon \,|\, Z_1,\, Z_2) = 0$ and $E(\varepsilon) = 0$. However, $E(\varepsilon \,|\, X) \neq 0$ since ε and ξ are correlated. To verify our asymptotic theory, we generate the sample with three different sample sizes: $n = 100$, 250, and 500. For each sample size, we replicate the design 500 times. We first report the mean and standard deviation of constant coefficient estimators among 500 replications for each sample size. For functional coefficient estimators, the MADE is computed for each function and each sample size, respectively.

Table 2 displays the estimation of $\boldsymbol{\beta}$ for the three-stage estimation corresponding to the estimator (6) with correcting for endogeneity, and for the standard local estimation corresponding to procedures proposed by Zhang *et al.* (2002) for ordinary semi-varying coefficient models without correcting for endogeneity. Both estimators give a similar asymptotically unbiased estimation for $\widehat{\beta}_1$ because there is no endogeneity involved in this parameter. As

Table 2: The Mean and Standard Deviation of the Constant Coefficients Estimates in the Simulated Example

Sample Size	Statistics	Three-Stage Estimation		Standard Estimation	
		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
100	Mean	0.9944	-0.9556	0.9900	-0.8843
	Std. Dev.[a]	0.0492	0.0571	0.1315	0.0599
250	Mean	0.9997	-0.9859	1.0002	-0.8747
	Std. Dev.	0.0342	0.0326	0.0911	0.0336
500	Mean	1.0000	-0.9989	0.9999	-0.8714
	Std. Dev.	0.0292	0.0159	0.0443	0.0197

[a]Standard deviation of 500 replications.

the sample size increases, both of them converge. However, a totally different phenomenon exhibits for $\widehat{\beta}_2$ (the estimator of constant coefficient in front of the endogenous variable X_2). Although both methods have similar standard deviations, a large bias is resulted from the standard local linear estimation. On the other hand, the three-stage estimator is asymptotically unbiased.

Figure 2 depicts boxplots of the 500 MADE values of the functional coefficient estimators using both three-stage local linear method in Figure 2(a) and ordinary local linear method without the adjustment for endogeneity in Figure 2(b), respectively. From Figure 2(a), we find that, for functional coefficients $g_1(\cdot)$, $g_2(\cdot)$ and $g_3(\cdot)$, all MADE values of the three-stage local linear estimator converge toward zero. But this is not true for ordinary local linear method presented in Figure 2(b). Clearly, $g_2(\cdot)$ is the only functional coefficient with the MADE value convergent toward zero because it is a coefficient of the exogenous variable Z_2. Note that $g_1(\cdot)$ is not directly involved in endogeneity, but the previous step bias of $\widehat{\beta}_2$ (the estimator of constant coefficient in front of the endogenous variable X_2), is carried over to the third stage and resides in $g_1(\cdot)$. The MADE value of $g_1(\cdot)$ basically converges to a positive constant. Therefore, the estimate of $g_1(\cdot)$ is biased. This is different from functional coefficient IV model of CDXW (2006) in which $g_1(\cdot)$ is not biased. Finally, the estimate of $g_3(\cdot)$ is also biased because of ignoring endogeneity.

Next, a typical example with the sample size $n = 500$ is chosen to present the proposed constant and functional coefficient estimators, as well as associated covariance estimators. The typical sample is chosen based on its MADE value equal to the median of the 500 MADE values. Table 3 lists the estimated values of $\boldsymbol{\beta}$ and their corresponding estimated

Table 3: A Sample Estimation of the Constant Coefficients in the Simulated Example

Parameters	$\widehat{\beta}_1$	$\widehat{\beta}_2$
Estimated Value	0.9825	-0.9718
Est. Std. Dev.[a]	0.0964	0.0453

[a]Estimated Standard Deviation

standard deviation. The theoretical values of them are 0.0447 and 0.0213, respectively. Our estimated values are a little bit over the them, which is sort expected because of the additional variation introduced by the over smooth at the first and second stage. But they are still on the ballpark. This shows the feasibility of our proposed estimator for both the constant coefficients and corresponding covariance structures.

Finally, Figure 3(a), (b), and (c) show the functions $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ (solid lines), their three-stage local linear estimators (dashed lines) with the corresponding 95% pointwise confidence intervals (dotted lines), and the ordinary local linear estimators that do not adjust for endogeneity (dashed-dotted lines). From Figure 3(a), we find that our estimator of $g_1(\cdot)$ is slightly better than the estimator without correcting endogeneity. Although estimation of $g_1(\cdot)$ does not involve endogenous variable, as explained above, it carries the bias of $\widehat{\beta}_2$, which is the coefficient of the endogenous variable X_2. From Figure 3(b), one can observe that both estimators fit closely to the true function $g_2(\cdot)$ since it does not involve endogeneity. From Figure 3(c), one can see a big difference between two estimators. The ordinary estimator is always over-estimated by ignoring the correlation between endogenous variables X_1 and X_2, and the random error ε. However, our proposed estimator fits this function well. Also note that the true function always stays within the 95% pointwise confidence intervals of our proposed estimator. That indicates that the finite sample performance of both the functional estimators and the estimators for their asymptotic variances is fairly good.

6 Proofs of Theorems

Throughout this section and the Appendix, we use the same notation as used in previous sections and we denote by C a generic constant, which may take different values at different appearances. Note that some proofs of Theorems 1-4 are similar to those in CDXW (2006) so that they are presented here briefly. To prove Theorems 1-4, the following lemmas are needed but their detailed proofs are given in the Appendix. First, the following lemma, due to Mack and Silverman (1982), is stated here without proof.

Lemma 1 . *Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i's are scalar random variables. Assume further that $E|y|^s < \infty$ and $\sup_x \int |y|^s f(x,y)dy < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support and satisfying a Lipschitz condition. Then,*

$$\sup_{x \in D} \left| \frac{1}{n} \sum_{i=1}^{n} \left[K_h(X_i - x)Y_i - E\left\{K_h(X_i - x)Y_i\right\}\right] \right| = O_p\left(\left(\frac{\log(1/h)}{nh^{d_x}} \right)^{1/2} \right),$$

provided that $n^{2\varepsilon-1} h^{d_x} \to \infty$ for some $\varepsilon < 1 - s^{-1}$ and d_x is the dimension of X.

As a consequence of Lemma 1, the following results are obtained.

Lemma 2 . *Under Assumptions 1-5, uniformly in $\mathbf{z}_{11}$, then,*

$$\frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W} \widehat{\bar{\mathbf{\Pi}}} = f_{11}(\mathbf{z}_{11})\, \mathbf{\Omega}_0(\mathbf{z}_{11}) + o_p(1) \quad \text{and} \quad \frac{1}{n} \widehat{\mathbf{\Pi}}^T \widetilde{\mathbf{W}} \widehat{\mathbf{\Pi}} = f_{11}(\mathbf{z}_{11})\, \mathbf{H}_3\, \tilde{\mathbf{\Omega}}_0(\mathbf{z}_{11})\, \mathbf{H}_3 + o_p(1),$$

where $\mathbf{H}_3 = diag\{1, \ldots, 1, h_3, \ldots, h_3\}$ is a $(d+l)(k+1) \times (d+l)(k+1)$ matrix with the first $(d+l)$ diagonal elements being 1's and the rest diagonal elements h_3's.

Now, we embark on establishing Theorems 1 and 2. First, $\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p$ is decomposed into three terms as follows:

$$\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p \equiv \mathbf{P} + \mathbf{Q} + \mathbf{R}, \tag{11}$$

where with $\mathbf{G} = (E(Y_1|\boldsymbol{\mathcal{Z}}_1), \ldots, E(Y_n|\boldsymbol{\mathcal{Z}}_n))^T$,

$$\mathbf{P} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}} \right)^{-1} \frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \left\{ \mathbf{Y} - \mathbf{G} \right\},$$

$$\mathbf{Q} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}} \right)^{-1} \frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \left\{ \mathbf{G} - \bar{\mathbf{\Pi}}\, \mathbf{\Theta}(\mathbf{Z}_{i11}) \right\},$$

and

$$\mathbf{R} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}} \right)^{-1} \frac{1}{n} \widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i (\bar{\mathbf{\Pi}} - \widehat{\bar{\mathbf{\Pi}}})\, \mathbf{\Theta}(\mathbf{Z}_{i11}).$$

Next, it is shown that $\mathbf{R}$ and $\mathbf{P}$ contribute to the first and second step variances respectively with the convergence rate $1/n$ instead of $1/(nh_2^m)$ in CDXW (2006) for nonparametric estimators, and that $\mathbf{Q}$ consists of the bias. First, consider $\mathbf{P}$, the first term on the right hand side of (11), which can be decomposed into two terms as follows:

$$\mathbf{P} \equiv \mathbf{P}_1 + \mathbf{P}_2,$$

where

$$\mathbf{P}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}\sum_{s=1}^{n} \mathbf{\Pi}_{*,s}\, L_{h_2}(\mathbf{Z}_{s11} - \mathbf{Z}_{i11})\, \eta_s,$$

and

$$\mathbf{P}_2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}\sum_{s=1}^{n} \begin{pmatrix} \widehat{\mathbf{\Pi}}_{1,-s} - \mathbf{\Pi}_{s1} \\ \widehat{\mathbf{\Pi}}_{2,-s} - \mathbf{\Pi}_{s2} \end{pmatrix} L_{h_2}(\mathbf{Z}_{s11} - \mathbf{Z}_{i11})\, \eta_s.$$

It is shown that $\mathbf{P}_1$ contributes to the second step variance from η but $\mathbf{P}_2$ is a higher order term comparing to $\mathbf{P}_1$, which are summarized in the following lemma with its proof given in the Appendix.

Lemma 3 . *Under Assumptions 1-6, then,*

$$n\ Var(\mathbf{P}_1) = \mathbf{e}^T E\left(\mathbf{\Omega}_0^{-1}\mathbf{\Omega}_\eta\mathbf{\Omega}_0^{-1}\right) \mathbf{e} + o(1) \quad \textit{and} \quad \mathbf{P}_2 = o_p\left(1/\sqrt{n}\right).$$

Next, consider the term $\mathbf{Q}$, the second term of (11). Here, the same decomposition as for $\mathbf{P}$ is employed. Define

$$\mathbf{Q} \equiv \mathbf{Q}_1 + \mathbf{Q}_2,$$

where

$$\mathbf{Q}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}\bar{\mathbf{\Pi}}^T \mathbf{W}_i \left\{\mathbf{G} - \bar{\mathbf{\Pi}}\,\mathbf{\Theta}(\mathbf{Z}_{i11})\right\}$$

and

$$\mathbf{Q}_2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}(\widehat{\bar{\mathbf{\Pi}}} - \bar{\mathbf{\Pi}})^T \mathbf{W}_i \left\{\mathbf{G} - \bar{\mathbf{\Pi}}\,\mathbf{\Theta}(\mathbf{Z}_{i11})\right\}.$$

Similar to Lemma 3, the following lemma is for $\mathbf{Q}$.

Lemma 4 . *Under Assumptions 1-8, then,*

$$\mathbf{Q}_1 = bias_p + o_p\left(h_2^2\right), \quad \textit{and} \quad \mathbf{Q}_2 = o_p\left(h_2^2\right).$$

Finally, it is shown that $\mathbf{R}$, the third term on the right hand side of (11), contributes to the bias and variance from the first step. Similar to the decompositions for $\mathbf{P}$ and $\mathbf{Q}$, $\mathbf{R}$ can be split as

$$\mathbf{R} \equiv \mathbf{R}_1 + \mathbf{R}_2,$$

where

$$\mathbf{R}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}\bar{\mathbf{\Pi}}^T \mathbf{W}_i(\bar{\mathbf{\Pi}} - \widehat{\bar{\mathbf{\Pi}}})\,\mathbf{\Theta}(\mathbf{Z}_{i11}),$$

and

$$\mathbf{R}_2 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}(\widehat{\bar{\mathbf{\Pi}}} - \bar{\mathbf{\Pi}})^T \mathbf{W}_i(\bar{\mathbf{\Pi}} - \widehat{\bar{\mathbf{\Pi}}})\,\mathbf{\Theta}(\mathbf{Z}_{i11}).$$

Further, $\mathbf{R}_1$ is split into bias and variance parts as

$$\mathbf{R}_1 \equiv \mathbf{R}_1^B + \mathbf{R}_1^V,$$

where

$$\begin{aligned}\mathbf{R}_1^B = &-\frac{1}{n}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \frac{1}{n}\sum_{1\le s\neq t\le n} \mathbf{\Pi}_{*,s}\, L_{h_2}(\mathbf{Z}_{s11} - \mathbf{Z}_{i11}) W_0^{h_1}(\boldsymbol{\mathcal{Z}}_t - \boldsymbol{\mathcal{Z}}_s)\\ &\times \left[\frac{1}{2}(\boldsymbol{\mathcal{Z}}_t - \boldsymbol{\mathcal{Z}}_s)^T \left\{\sum_{j=1}^{l} g_{d+j}(\mathbf{Z}_{i11})\,\pi_j''(\boldsymbol{\mathcal{Z}}_s) + \sum_{k=1}^{p} \beta_{m+k}\,\pi_{l+k}''(\boldsymbol{\mathcal{Z}}_s)\right\}(\boldsymbol{\mathcal{Z}}_t - \boldsymbol{\mathcal{Z}}_s) + o_p(h_1^2)\right]\end{aligned}$$

and

$$\mathbf{R}_1^V = \frac{-1}{n^2}\sum_{i=1}^{n} \mathbf{e}^T \left(\frac{1}{n}\widehat{\bar{\mathbf{\Pi}}}^T \mathbf{W}_i \widehat{\bar{\mathbf{\Pi}}}\right)^{-1} \sum_{1\le s\neq t\le n} \mathbf{\Pi}_{*,s}\, L_{h_2}(\mathbf{Z}_{s11} - \mathbf{Z}_{i11})\, W_0^{h_1}(\boldsymbol{\mathcal{Z}}_t - \boldsymbol{\mathcal{Z}}_s)\mathbf{\Theta}_*(\mathbf{Z}_{i11})^T \boldsymbol{\xi}_t.$$

Similar to $\mathbf{Q}$, it is shown in the following lemma that both $\mathbf{R}_1^B$ and $\mathbf{R}_2$ are a higher order and $\mathbf{R}_1^V$ contributes to the asymptotic variance.

Lemma 5 . *Under Assumptions 1-8, then,*

$$\mathbf{R}_1^B = O_p(h_1^2) = o_p(h_2^2), \;\; n\ Var(\mathbf{R}_1^V) \to \mathbf{e}^T E\left(\mathbf{\Omega}_0^{-1}\mathbf{\Omega}_\xi\mathbf{\Omega}_0^{-1}\right)\mathbf{e}, \; \textit{and}\;\; \mathbf{R}_2 = o_p(h_2^2) + o_p\left(n^{-1/2}\right).$$

We now proceed with the proofs of Theorems 1 and 2.

Proof of Theorem 1: Since $E(\mathbf{P}_1) = 0$ and $E(\mathbf{R}_1^V) = 0$, it follows from Lemmas 3 and 5 that

$$\mathbf{P}_1 = O_p\,(1/\sqrt{n}), \quad \text{and} \quad \mathbf{R}_1^V = O_p\,(1/\sqrt{n}),$$

which, in conjunction with (11), the decompositions of $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$, and Lemmas 2 - 5, implies that

$$\begin{aligned}\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p &= \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{R}_1 + \mathbf{R}_2\\ &= O_p\,(1/\sqrt{n}) + o_p\,(1/\sqrt{n}) + bias_p + o_p\,(h_2^2) + o_p\,(h_2^2) + O_p\,(1/\sqrt{n}) + o_p\,(1/\sqrt{n})\\ &= O_p\,(1/\sqrt{n}) + bias_p + o_p\,(h_2^2). \qquad (12)\end{aligned}$$

This proves the theorem. $\square$

Proof of Theorem 2: By (12) and Lemmas 2 - 5, one has

$$\sqrt{n}\,\left[\widehat{\mathbf{\Theta}}_p - \mathbf{\Theta}_p - bias_p + o_p(h_2^2)\right] = \sqrt{n}\,\left\{\mathbf{P}_1 + \mathbf{R}_1^V\right\} + o_p\,(1).$$

To establish the theorem, it suffices to show that the right hand side is asymptotically normally distributed. To this end, by the definitions of $\mathbf{P}_1$ and $\mathbf{R}_1^V$ and Lemma 2, one has

$$\frac{1}{n}\widehat{\widehat{\mathbf{\Pi}}}^T\mathbf{W}_i\,\widehat{\widehat{\mathbf{\Pi}}} = \mathbf{\Omega}_0^{-1}(\mathbf{Z}_{i11}) + o_p(1).$$

Applying Lemma 1 to the sum of s in $\mathbf{P}_1$ and to the sum of s and t in $\mathbf{R}_1^V$, and combining these with the above equation, one has

$$\begin{aligned}
&\sqrt{n}\,\left\{\mathbf{P}_1 + \mathbf{R}_1^V\right\} \\
= &\ \frac{1}{\sqrt{n}}\left[\sum_{i=1}^n\left\{\mathbf{e}^T\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{i11})\mathbf{\Pi}_{*,i}\eta_i + o_p(1)\right\} - \sum_{i=1}^n\left\{\mathbf{e}^T\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{i11})\mathbf{\Pi}_{*,i}\mathbf{\Theta}_*(\mathbf{Z}_{i11})^T\boldsymbol{\xi}_i + o_p(1)\right\}\right] \\
= &\ \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\mathbf{e}^T\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{i11})\mathbf{\Pi}_{*,i}\left\{\eta_i - \mathbf{\Theta}_*(\mathbf{Z}_{i11})^T\,\boldsymbol{\xi}_i\right\}\right] + o_p(1).
\end{aligned}$$

By the central limit theorem, one can easily establish the asymptotical normality of the right hand side of the above equation, omitted. $\square$

Proofs of Theorems 3 and 4: First, observe that

$$\widehat{g}_j(\mathbf{z}_{11}) - g_j(\mathbf{z}_{11}) \equiv \mathbf{e}_j^T\,\mathbf{H}_3\,\left(n^{-1}\,\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,\widehat{\mathbf{\Pi}}\right)^{-1}\left\{\widetilde{\mathbf{P}} + \widetilde{\mathbf{Q}} + \widetilde{\mathbf{R}} + \tilde{\mathbf{a}}\right\}, \tag{13}$$

where $\widetilde{\mathbf{G}} = (\mathbf{\Pi}_{11}^T\widetilde{\mathbf{\Theta}}(\mathbf{Z}_{111}),\cdots,\mathbf{\Pi}_{n1}^T\widetilde{\mathbf{\Theta}}(\mathbf{Z}_{n11}))^T$, $\widetilde{\mathbf{Q}} = \frac{1}{n}\,\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,(\widetilde{\mathbf{G}} - \mathbf{\Pi}\,\widetilde{\mathbf{\Theta}})$, $\widetilde{\mathbf{R}} = \frac{1}{n}\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,(\mathbf{\Pi} - \widehat{\mathbf{\Pi}})\,\widetilde{\mathbf{\Theta}}$, $\widetilde{\mathbf{P}} = \frac{1}{n}\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\left\{\mathbf{Y} - \mathbf{T}_2\mathbf{\Theta}_p - \widetilde{\mathbf{G}}\right\} = \frac{1}{n}\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,\{\mathbf{Y} - \mathbf{G}\}$, $\tilde{\mathbf{a}} = \frac{1}{n}\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,\left\{\mathbf{T}_2\,(\mathbf{\Theta}_p - \widehat{\mathbf{\Theta}}_p)\right\}$, and $\mathbf{T}_2 = \begin{pmatrix}\mathbf{Z}_{113} & \ldots & \mathbf{Z}_{n13}\\ \mathbf{X}_{12} & \ldots & \mathbf{X}_{n2}\end{pmatrix}^T$. We now show that the proofs of Theorems 3 and 4 can be established by following those for Theorems 2 and 3 in CDXW (2006). To this end, we compare (13) with the corresponding expression in CDXW (2006) and find that the only difference is the additional term $\tilde{\mathbf{a}}$. It is shown in the Appendix that $\tilde{\mathbf{a}}$ is asymptotically negligible; that is

$$\mathbf{e}_j^T\,\mathbf{H}_3\,\left(n^{-1}\,\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,\widehat{\mathbf{\Pi}}\right)^{-1}\tilde{\mathbf{a}} = o_p\,(h_3^2) + o_p\,\left((n\,h_3^k)^{-1/2}\right). \tag{14}$$

Therefore, the estimator of the functional coefficients of partially varying coefficient IV models has the same asymptotic properties as in the functional coefficient IV models in CDXW (2006). Hence, the proofs of Theorems 3 and 4 are the same as those for Theorems 2 and 3 in CDXW (2006). The detailed proofs are referred to CDXW (2006) and omitted. $\square$

Appendix

Proof of Lemma 2: Observe that

$$\begin{aligned}
\frac{1}{n}\,\hat{\bar{\mathbf{\Pi}}}^T\mathbf{W}\,\hat{\bar{\mathbf{\Pi}}} &= n^{-1}\,\bar{\mathbf{\Pi}}^T\mathbf{W}\,\bar{\mathbf{\Pi}} + n^{-1}\,\bar{\mathbf{\Pi}}^T\mathbf{W}\,(\hat{\bar{\mathbf{\Pi}}}-\bar{\mathbf{\Pi}}) + n^{-1}\,(\hat{\bar{\mathbf{\Pi}}}-\bar{\mathbf{\Pi}})^T\mathbf{W}\,\bar{\mathbf{\Pi}} \\
&\quad + n^{-1}\,(\hat{\bar{\mathbf{\Pi}}}-\bar{\mathbf{\Pi}})^T\mathbf{W}\,(\hat{\bar{\mathbf{\Pi}}}-\bar{\mathbf{\Pi}}) \\
&\equiv I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

We first look at

$$I_1 = \frac{1}{n}\sum_i L_{h_2}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,\mathbf{\Pi}_{*,\,i}\mathbf{\Pi}_{*,\,i}^T.$$

By Lemma 1, uniformly in $\mathbf{z}_{11}$,

$$I_1 = f_{11}(\mathbf{z}_{11})\,\mathbf{\Omega}_0(\mathbf{z}_{11})\, + o_p(1).$$

Next, we consider

$$I_1 = \frac{1}{n}\sum_i L_{h_2}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,\mathbf{\Pi}_{*,\,i}\left(\widehat{\mathbf{\Pi}}_{*,\,i}-\mathbf{\Pi}_{*,\,i}\right)^T.$$

Applying Lemma 1 to elements of $\widehat{\mathbf{\Pi}}_{*,\,i}$ and the sum i, we have

$$\begin{aligned}
I_2 &= \frac{1}{n}\sum_i L_{h_2}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,\mathbf{\Pi}_{*,\,i}\,\mathbf{1}^T o_p\,(1) \\
&= o_p\,(1)\left\{E\left(\mathbf{\Pi}_*\mathbf{1}^T|\mathbf{Z}_{11}=\mathbf{z}_{11}\right)+o_p\,(1)\right\} = o_p(1),
\end{aligned}$$

where $\mathbf{1}$ is a vector of 1's with the same dimension as $\mathbf{\Pi}_{*,\,i}$. Similarly, it can be shown that $I_3 = o_p(1)$ and $I_4 = o_p(1)$. This proves the lemma. □

Proof of Lemma 3: By Lemma 2, one has

$$\begin{aligned}
\mathbf{P}_1 &= \frac{1}{n}\sum_{i=1}^n \mathbf{e}^T\,[f_{11}(\mathbf{Z}_{i11})\,\mathbf{\Omega}_0(\mathbf{Z}_{i11})\, + o_p(1)]^{-1}\,\frac{1}{n}\sum_{s=1}^n \mathbf{\Pi}_{*,s}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\,\eta_s \\
&= \frac{1}{n^2}\sum_{s=1}^n\sum_{i=1}^n \mathbf{e}^T\,[f_{11}(\mathbf{Z}_{i11})\,\mathbf{\Omega}_0(\mathbf{Z}_{i11})\, + o_p(1)]^{-1}\,\mathbf{\Pi}_{*,s}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\,\eta_s.
\end{aligned}$$

An application of Lemma 1 to the sum of i gives

$$\mathbf{P}_1 = \frac{1}{n}\sum_{s=1}^n\left\{\mathbf{e}^T\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{s11})\mathbf{\Pi}_{*,s}\,\eta_s + o_p(1)\right\}.$$

Then,

$$\begin{aligned}
n\,\mathrm{Var}(\mathbf{P}_1) &= \frac{1}{n}\sum_{s=1}^n E\left[\mathbf{e}^T\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{s11})\mathbf{\Pi}_{*,s}\,E(\eta_s^2|\boldsymbol{\mathcal{Z}}_s)\,\mathbf{\Pi}_{*,s}^T\,\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{s11})\mathbf{e}+o_p\,(1)\right] \\
&= \mathbf{e}^T\,E\left[\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{11})\,\mathbf{\Omega}_\pi(\boldsymbol{\mathcal{Z}})\,\sigma_\eta^2(\boldsymbol{\mathcal{Z}})\mathbf{\Omega}_0^{-1}(\mathbf{Z}_{11})+o_p\,(1)\right]\,\mathbf{e} \\
&= \mathbf{e}^T\,E\left\{\mathbf{\Omega}_0^{-1}\mathbf{\Omega}_\eta\mathbf{\Omega}_0^{-1}\right\}\,\mathbf{e}+o\,(1)
\end{aligned}$$

Thus, the first part of this lemma is proved. To show the second part, by Lemma 1 and 2, we have

$$\begin{aligned}
\mathbf{P}_2 &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{e}^T\left[f_{11}(\mathbf{Z}_{i11})\,\boldsymbol{\Omega}_0(\mathbf{Z}_{i11})\,+o_p(1)\right]^{-1}\frac{1}{n}\sum_{s=1}^{n}\mathbf{1}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\,\eta_s\,o_p\,(1)\\
&= \frac{1}{n}\sum_{s=1}^{n}\left\{\mathbf{e}^T\boldsymbol{\Omega}_0^{-1}(\mathbf{Z}_{s11})\mathbf{1}\,\eta_s+o_p(1)\right\}\,o_p\,(1).
\end{aligned}$$

Similar to the calculation for $\mathbf{P}_1$, we have

$$n\,E(\mathbf{P}_2\mathbf{P}_2^T)=\mathbf{e}^T\;E\left\{\boldsymbol{\Omega}_0^{-1}\mathbf{1}\mathbf{1}^T\boldsymbol{\Omega}_0^{-1}\sigma_\eta^2\right\}\mathbf{e}\,o(1)=o\,(1).$$

Thus $\mathbf{P}_2=o_p\,(1/\sqrt{n})$. □

Proof of Lemma 4: By Lemma 2, one has

$$\begin{aligned}
&\mathbf{Q}_1\\
=\;& \frac{1}{n^2}\sum_{s=1}^{n}\sum_{i=1}^{n}\mathbf{e}^T\left\{f_{11}(\mathbf{Z}_{i11})\boldsymbol{\Omega}_0(\mathbf{Z}_{i11})+o_p(1)\right\}^{-1}\,\boldsymbol{\Omega}_\pi(\boldsymbol{\mathcal{Z}}_s)\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\\
\times\;& \begin{pmatrix}
(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})^Tg_1'(\mathbf{Z}_{i11})+\frac{1}{2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})^Tg_1''(\mathbf{Z}_{i11})(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\\
\vdots\\
(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})^Tg_{d+l}'(\mathbf{Z}_{i11})+\frac{1}{2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})^Tg_{d+l}''(\mathbf{Z}_{i11})(\mathbf{Z}_{i11})(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\\
\mathbf{0}_{(m+p)\times 1}
\end{pmatrix}+o_p(h_2^2).
\end{aligned}$$

Similar to $\mathbf{P}_1$, by applying Lemma 1 to sum of i, we obtain

$$\begin{aligned}
\mathbf{Q}_1 &= \frac{h_2^2}{2\,n}\sum_{s=1}^{n}\mathbf{e}^T\left\{\boldsymbol{\Omega}_0(\mathbf{Z}_{s11})+o_p(1)\right\}^{-1}\,\boldsymbol{\Omega}_\pi(\boldsymbol{\mathcal{Z}}_s)\\
&\quad\times\begin{pmatrix}
\operatorname{tr}\left[\boldsymbol{\mu}_2(L)\left\{2\,g_1'\,{f_{11}'}^T/f_{11}+g_1''\right\}\right]\\
\vdots\\
\operatorname{tr}\left[\boldsymbol{\mu}_2(L)\left\{2\,g_{d+l}'\,{f_{11}'}^T/f_{11}+g_{d+l}''\right\}\right]\\
\mathbf{0}_{(m+p)\times 1}
\end{pmatrix}+o_p(h_2^2)\\
&= \frac{h_2^2}{2}\,\mathbf{e}^T\,E\left\{\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Omega}_\pi(\boldsymbol{\mathcal{Z}})\begin{pmatrix}
\operatorname{tr}\left[\boldsymbol{\mu}_2(L)\left\{2\,g_1'\,{f_{11}'}^T/f_{11}+g_1''\right\}\right]\\
\vdots\\
\operatorname{tr}\left[\boldsymbol{\mu}_2(L)\left\{2\,g_{d+l}'\,{f_{11}'}^T/f_{11}+g_{d+l}''\right\}\right]\\
\mathbf{0}_{(m+p)\times 1}
\end{pmatrix}\right\}+o_p\,(h_2^2).
\end{aligned}$$

This proves the first part of the lemma. Similar to Lemma A3 in CDXW (2006), one can show easily that $\mathbf{Q}_2=o_p\,(h_2^2)$. Therefore, this lemma is established. □

Proof of Lemma 5: By Lemma 2 again,

$$\begin{aligned}
&\mathbf{R}_1^B\\
=\;& -\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}^T\left[f_{11}(\mathbf{Z}_{i11})\boldsymbol{\Omega}_0(\mathbf{Z}_{i11})\,+o_p(1)\right]^{-1}\frac{1}{n}\sum_{s=1}^{n}\sum_{t\neq s}\boldsymbol{\Pi}_{*,s}L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})W_0^{h_1}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)\\
&\times\left[\frac{1}{2}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)^T\left\{\sum_{j=1}^{l}g_{d+j}(\mathbf{Z}_{i11})\pi_j''(\boldsymbol{\mathcal{Z}}_s)+\sum_{k=1}^{p}\beta_{m+k}\pi_{l+k}''(\boldsymbol{\mathcal{Z}}_s)\right\}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)+o_p(h_1^2)\right]
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{h_1^2}{2}\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}^T\left[f_{11}(\mathbf{Z}_{i11})E\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\mathbf{Z}_{i11}\right\} + o_p(1)\right]^{-1} \\
&\quad \times\frac{1}{n}\sum_{s=1}^{n}\sum_{t\neq s}\boldsymbol{\Pi}_{*,s}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\frac{1}{(n-1)f(\boldsymbol{\mathcal{Z}}_s)}K_{h_1}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s) \\
&\quad \times\left[(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)^T\left\{\sum_{j=1}^{l}g_{d+j}(\mathbf{Z}_{i11})\pi_j''(\boldsymbol{\mathcal{Z}}_s)+\sum_{k=1}^{p}\beta_{m+k}\pi_{l+k}''(\boldsymbol{\mathcal{Z}}_s)\right\}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)+o_p(h_1^2)\right].
\end{aligned}
$$

By applying Lemma 1,

$$
\begin{aligned}
\mathbf{R}_1^B &= -\frac{h_1^2}{2}\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}^T\left[f_{11}(\mathbf{Z}_{i11})\,E\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\,\mathbf{Z}_{i11}\right\} + o_p(1)\right]^{-1}\frac{1}{n}\sum_{s=1}^{n}\boldsymbol{\Pi}_{*,s}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11}) \\
&\quad\times\left[\sum_{j=1}^{l}g_{d+j}(\mathbf{Z}_{i11})\mathrm{tr}\left\{\boldsymbol{\mu}_2(K)\pi_j''\right\}+\sum_{k=1}^{p}\beta_{m+k}\mathrm{tr}\left\{\boldsymbol{\mu}_2(K)\pi_{l+k}''\right\}\right]+o_p\left(h_1^2\right) \\
&= -\frac{h_1^2}{2}\frac{1}{n}\sum_{s=1}^{n}\mathbf{e}^TE^{-1}\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\,\mathbf{Z}_{s11}\right\}\boldsymbol{\Pi}_{*,s}\,\boldsymbol{\Theta}_*^T\begin{pmatrix}\mathrm{tr}\left\{\boldsymbol{\mu}_2(K)\pi_1''\right\}\\ \vdots \\ \mathrm{tr}\left\{\boldsymbol{\mu}_2(K)\pi_{l+p}''\right\}\end{pmatrix}+o_p\left(h_1^2\right) \\
&= O_p\left(h_1^2\right)=o_p\left(h_2^2\right).
\end{aligned}
$$

This proves the first assertion of the lemma. Next, we prove the second result of the lemma. By Lemma 2 again,

$$
\begin{aligned}
\mathbf{R}_1^V &= -\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}^T\left[f_{11}(\mathbf{Z}_{i11})\,\boldsymbol{\Omega}_0(\mathbf{Z}_{i11}) + o_p(1)\right]^{-1} \\
&\quad \frac{1}{n}\sum_{s=1}^{n}\sum_{t\neq s}\boldsymbol{\Pi}_{*,s}\,L_{h_2}(\mathbf{Z}_{s11}-\mathbf{Z}_{i11})\frac{1}{(n-1)\,f(\boldsymbol{\mathcal{Z}}_s)}K_{h_1}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)\boldsymbol{\Theta}_*(\mathbf{Z}_{i11})^T\boldsymbol{\xi}_t \\
&= -\frac{1}{n}\sum_{s=1}^{n}\sum_{t\neq s}\mathbf{e}^T\left[\boldsymbol{\Omega}_0(\mathbf{Z}_{s11})+o_p(1)\right]^{-1}\boldsymbol{\Pi}_{*,s}\frac{1}{(n-1)\,f(\boldsymbol{\mathcal{Z}}_s)}K_{h_1}(\boldsymbol{\mathcal{Z}}_t-\boldsymbol{\mathcal{Z}}_s)\,\boldsymbol{\Theta}_*(\mathbf{Z}_{s11})^T\boldsymbol{\xi}_t \\
&= -\frac{1}{n}\sum_{s=1}^{n}\mathbf{e}^T\left[E\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\,\mathbf{Z}_{s11}\right\}+o_p(1)\right]^{-1}\boldsymbol{\Pi}_{*,s}\,\boldsymbol{\Theta}_*(\mathbf{Z}_{i11})^T\boldsymbol{\xi}_s.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
n\,\mathrm{Var}(\mathbf{R}_1^V) &= \frac{1}{n}\sum_{t=1}^{n}E\Big\{\mathbf{e}^T\left[E\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\,\mathbf{Z}_{t11}\right\}+o_p(1)\right]^{-1}\boldsymbol{\Omega}_\pi(\boldsymbol{\mathcal{Z}}_t)\,\boldsymbol{\Theta}_*^T\,\boldsymbol{\Sigma}_\xi\,\boldsymbol{\Theta}_* \\
&\quad\times\left[E\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\,\mathbf{Z}_{t11}\right\}+o_p(1)\right]^{-1}\mathbf{e}\Big\} \\
&= \mathbf{e}^T\,E\left[E^{-1}\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\mathbf{Z}_{11}\right\}\boldsymbol{\Omega}_\pi(\boldsymbol{\mathcal{Z}})\,\boldsymbol{\Theta}_*^T\,\boldsymbol{\Sigma}_\xi\,\boldsymbol{\Theta}_*\,E^{-1}\left\{\boldsymbol{\Omega}_0(\boldsymbol{\mathcal{Z}})|\mathbf{Z}_{11}\right\}+o_p\left(1\right)\right]\mathbf{e} \\
&= \mathbf{e}^T\,E\left\{\boldsymbol{\Omega}_0^{-1}\,\boldsymbol{\Omega}_\xi(\boldsymbol{\mathcal{Z}})\,\boldsymbol{\Omega}_0^{-1}\right\}\mathbf{e}+o\left(1\right).
\end{aligned}
$$

Finally, the proof of the third conclusion is similar to that for Lemma A3 in CDXW (2006), omitted. Therefore, this proves the lemma. □

Proof of (14): Similar to previous lemmas, $\widetilde{\mathbf{a}}$ is split as

$$
\widetilde{\mathbf{a}}\equiv\widetilde{\mathbf{a}}_1+\widetilde{\mathbf{a}}_2,
$$

where $\widetilde{\mathbf{a}}_1 = \frac{1}{n}\mathbf{\Pi}^T\widetilde{\mathbf{W}}\left\{\mathbf{T}_2\left(\mathbf{\Theta}_p - \widehat{\mathbf{\Theta}}_p\right)\right\}$ and $\widetilde{\mathbf{a}}_2 = \frac{1}{n}\left\{\widehat{\mathbf{\Pi}} - \mathbf{\Pi}\right\}^T\widetilde{\mathbf{W}}\left\{\mathbf{T}_2\left(\mathbf{\Theta}_p - \widehat{\mathbf{\Theta}}_p\right)\right\}$. First, investigate $\widetilde{\mathbf{a}}_1$. To this end, is further decomposed as follows

$$
\begin{aligned}
& \mathbf{H}_3^{-1}\widetilde{\mathbf{a}}_1 \\
= & \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{\Pi}_{i1} \\ \mathbf{\Pi}_{i1}\otimes(\mathbf{Z}_{i11}-\mathbf{z}_{11})/h_3\end{pmatrix} M_{h_3}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,(\mathbf{Z}_{i13}^T\mathbf{1}+\mathbf{X}_{i2}^T\mathbf{1})\,\left\{O_p\left(h_2^2\right)+O_p\left(1/\sqrt{n}\right)\right\} \\
\equiv & \widetilde{\mathbf{a}}_1^B + \widetilde{\mathbf{a}}_1^V,
\end{aligned}
$$

where

$$
\widetilde{\mathbf{a}}_1^B = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{\Pi}_{i1} \\ \mathbf{\Pi}_{i1}\otimes(\mathbf{Z}_{i11}-\mathbf{z}_{11})/h_3\end{pmatrix} M_{h_3}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,(\mathbf{Z}_{i13}^T\mathbf{1}+\mathbf{X}_{i2}^T\mathbf{1})\,O_p\left(h_2^2\right),
$$

and

$$
\widetilde{\mathbf{a}}_1^V = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{\Pi}_{i1} \\ \mathbf{\Pi}_{i1}\otimes(\mathbf{Z}_{i11}-\mathbf{z}_{11})/h_3\end{pmatrix} M_{h_3}(\mathbf{Z}_{i11}-\mathbf{z}_{11})\,(\mathbf{Z}_{i13}^T\mathbf{1}+\mathbf{X}_{i2}^T\mathbf{1})\,O_p\left(1/\sqrt{n}\right).
$$

Next, evaluate $\widetilde{\mathbf{a}}_1^B(s)$, the s-th element of $\widetilde{\mathbf{a}}_1^B$. For $1 \le s \le (d+l)$,

$$
\begin{aligned}
E\left\{\widetilde{\mathbf{a}}_1^B(s)\right\}^2 &= \Bigg(\frac{1}{n}E\left[\pi_{s1}^2(\boldsymbol{\mathcal{Z}}_1)\,M_{h_3}^2(\mathbf{Z}_{111}-\mathbf{z}_{11})(\mathbf{Z}_{113}^T\mathbf{1}+\mathbf{X}_{12}^T\mathbf{1})^2\right] \\
&\quad + E\left[\pi_{s1}(\boldsymbol{\mathcal{Z}}_1)\pi_{s1}(\boldsymbol{\mathcal{Z}}_2)\,M_{h_3}(\mathbf{Z}_{111}-\mathbf{z}_{11})M_{h_3}(\mathbf{Z}_{211}-\mathbf{z}_{11})\right. \\
&\quad \left.\times(\mathbf{Z}_{113}^T\mathbf{1}+\mathbf{X}_{12}^T\mathbf{1})(\mathbf{Z}_{213}^T\mathbf{1}+\mathbf{X}_{22}^T\mathbf{1})\right]\Bigg)\,O\left(h_2^4\right) \\
&= \left[O\left\{1/(n\,h_3)\right\}+O\left(1\right)\right]\,O\left(h_2^4\right) = O\left(h_2^4\right).
\end{aligned}
$$

Similarly, for $(d+l+1) \le s \le (d+l)\times(m+1)$, $E\left\{\widetilde{\mathbf{a}}_1^B(s)\right\}^2 = O\left(h_2^4\right)$ and $\widetilde{\mathbf{a}}_1^B = O_p(h_2^2)$. For the same reason, it can be shown that $\widetilde{\mathbf{a}}_1^V = O_p(1/\sqrt{n})$. Therefore, by Lemma 2,

$$
\begin{aligned}
\mathbf{e}_j^T\,\mathbf{H}_3\left(n^{-1}\,\widehat{\mathbf{\Pi}}^T\widetilde{\mathbf{W}}\,\widehat{\mathbf{\Pi}}\right)^{-1}\widetilde{\mathbf{a}} &= \mathbf{e}_j^T\left\{f_{11}(\mathbf{z}_{11})\,\widetilde{\mathbf{\Omega}}_0(\mathbf{z}_{11})\,+o_p\left(1\right)\right\}^{-1}\left\{\widetilde{\mathbf{a}}_1^B+\widetilde{\mathbf{a}}_1^V\right\} \\
&= O_p\left(h_2^2\right)+O_p\left(1/\sqrt{n}\right) = o_p\left(h_3^2\right)+o_p\left(1/\sqrt{n\,h_3^k}\right).
\end{aligned}
$$

This completes the proof of (14). □

References

Ahmad, I., S. Leelahanon and Q. Li (2005), Efficient estimation of a semiparametric partially linear varying coefficient Model. *Annals of Statistics* **33**, 258-283.

Ai, C. and X. Chen (2003), Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* **71**, 1795-1843.

Andrews, D.W.K. (1994), Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* **62**, 43-72.

Blundell, R. and J. Powell (2003), Endogeneity in nonparametric and semiparametric regression models. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. II (M. Dewatripont, L.P. Hansen and S.J. Turnovsky, eds.), Cambridge: Cambridge University Press.

Cai, Z. (2002a), Two-step likelihood estimation procedure for varying-coefficient models. *Journal of Multivariate Analysis* **81**, 189-209.

Cai, Z. (2002b), A two-stage approach to additive time series models. *Statistica Neerlandica* **56**, 415-433.

Cai, Z., M. Das, H. Xiong and X. Wu (2006), Functional coefficient instrumental variables models. *Journal of Econometrics* **133**, 207-241.

Card, D. (2001), Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* **69**, 1127-1160.

Carroll, R.J., D. Ruppert, T. Tosteson, C. Crainiceanu and M. Karagas (2004), Instrumental variables and nonparametric regression. *Journal of the American Statistical Association* **99**, 736-750.

Darolles, S., J.-P. Florens and E. Renault (2000), Non-parametric instrumental regression. *Mimeo*, GREMAQ, University of Toulouse.

Das, M. (2005), Instrumental variables estimators for nonparametric models with discrete endogenous regressors. *Journal of Econometrics* **124**, 335-361.

Das, M, W. Newey and F. Vella (2003), Nonparametric estimation of sample selection models. *The Review of Economic Studies* **70**, 33-58.

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

Hall, P. and J.L. Horowitz (2005), Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* **33**, 2904-2929.

Mack, Y.P. and B.W. Silverman (1982), Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**, 405-415.

Newey, W.K. and J.L. Powell (2003), Nonparametric instrumental variables estimation. *Econometrica* **71**, 1565-1578.

Newey, W.K., J.L. Powell and F. Vella (1999), Nonparametric estimation of triangular simultaneous equations models. *Econometrica* **67**, 565-603.

Pakes, A. and S. Olley (1995), A limit theorem for a smooth class of semiparametric estimators. *Journal of Econometrics* **65**, 295-332.

Park, S. (2003), Semiparametric instrumental variables estimation. *Journal of Econometrics* **112**, 381-399.

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

Schultz, T.P. (1997), *Human Capital, Schooling and Health*, IUSSP, XXIII, General Population Conference, Yale University.

Xiong, H. (2004), Some nonparametric and semiparametric instrumental variables models for economic data. *Ph.D. Dissertation*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.

Zhang, W., S.-Y. Lee and X. Song (2002), Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis* **82**, 166-188.

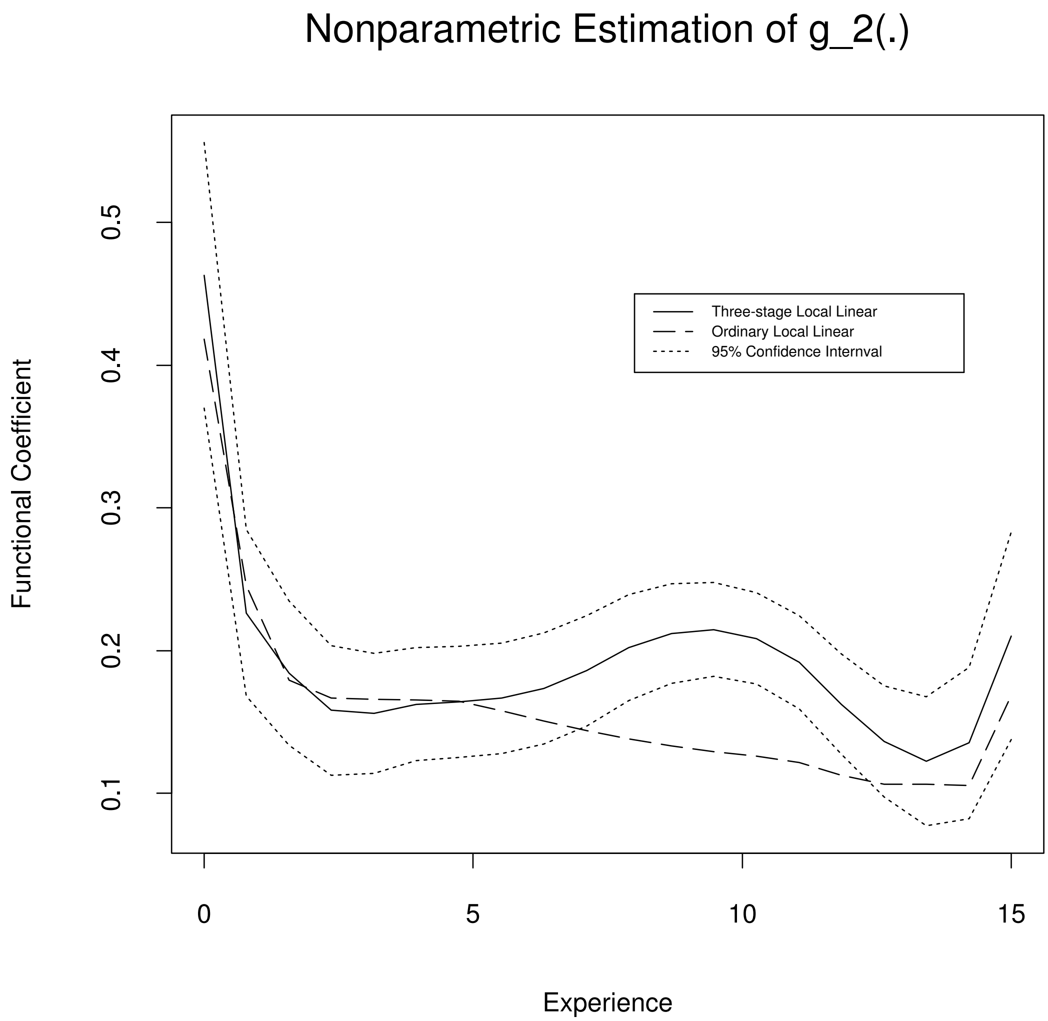
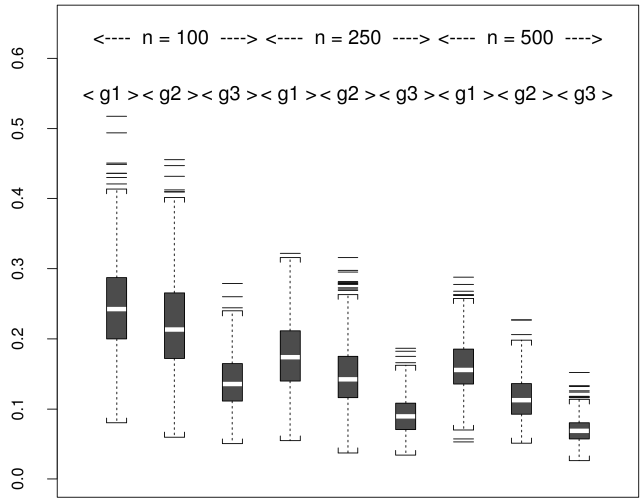


Figure 1: The Estimate of Functional Coefficient for Schooling in Structural Equation (10) The figure corresponds to the functional coefficient $g_2(\cdot)$, *graphing the three-stage local linear estimate (solid line) with point-wise* 95% *confidence intervals (dotted lines), and the ordinary nonparametric estimate (dashed line).*

(a) Boxplot of MADE Values (Three-stage Local Linear Method)



(b) Boxplot of MADE Values (Ordinary Local Linear Method)

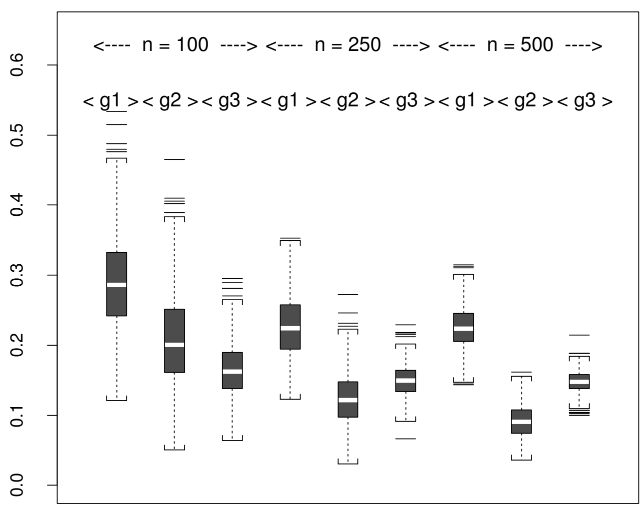


Figure 2: Simulation Results for the Simulated Example

Figures (a) and (b) give the boxplots of the 500 *MADE values in the estimation of* $g_1(z)$, $g_2(z)$, *and* $g_3(z)$, *respectively.*

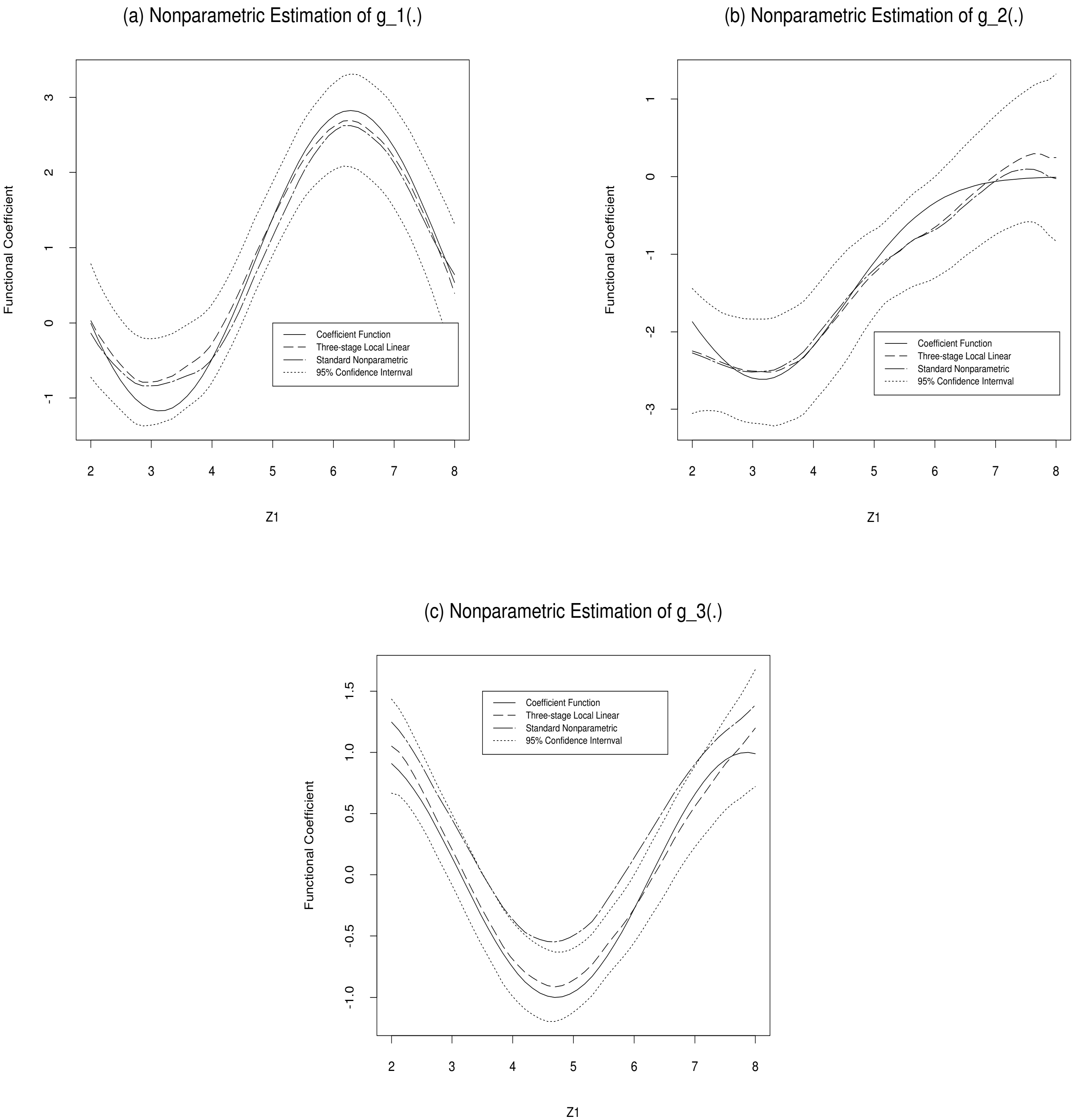


Figure 3: Simulation Results for the Simulated Example

Displayed in (a), (b), and (c) are the true coefficient functions (in solid line), the three-stage local linear estimators (dashed line) with the corresponding 95% point-wise confidence intervals (dotted lines), and the ordinary nonparametric estimator (dashed-dotted line).