Convergency and Divergency of Functional Coefficient Weak Instrumental Variables Models *

Zongwu Cai\textsuperscript{a,b} and Henong Li\textsuperscript{a}

\textsuperscript{a}Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
\textsuperscript{b}Wang Yanan Institute for Studies in Economics, Xiamen University, China

This Version: May 18, 2007

In this paper, we consider a simultaneous equations model under a functional coefficient representation for the structural equation of interest and adopt the local-to-zero assumptions as in Staiger and Stock (1997) and Hahn and Kuersteiner (2002) on the coefficients of the instruments in the reduced form equation. Under this functional coefficient representation, models are linear in endogenous components with coefficients governed by unknown functions of the predetermined exogenous variables. We propose a two-step estimation procedure to estimate the coefficient functions. The first step is to estimate a matrix of unknown parameters of the reduced form equation based on the least squares method, and the second step is to use the local linear fitting technique to estimate coefficient functions by using the estimated reduced forms as regressors. We investigate how the limiting distribution of the proposed nonparametric estimator changes as the parameterization is allowed for different degrees of weakness. As a result, our new theoretical findings are that the possible convergency of the proposed nonparametric estimators can be attained only for the nearly weak case and the rate of convergence for the nonparametric estimator for coefficient functions of endogenous variables is slower than the conventional rate. But the nonparametric estimator for coefficient functions of endogenous variables is divergent for both the weak and nearly non-identified cases. A Monte Carlo simulation is conducted to illustrate the finite sample performance of the resulting estimator and results support these theoretical findings.

KEY WORDS: Discontinuity; Divergence; Endogenous variables; Functional coefficient model; Weak instrumental variables; Local linear fitting; Simultaneous equations.

\textsuperscript{*}This research was supported, in part, by the National Science Foundation grant DMS-0404954 and funds provided by the University of North Carolina at Charlotte. Correspondence author: Tel. (704) 687-2650. Fax: (704) 687-6415. E-mail address: zcai@uncc.edu (Z. Cai).
1 Introduction

Since the seminal work by Staiger and Stock (1997), the literature has grown swiftly on the studies of weak instrumental variables (IV) models due to their various applications in economics and finance.\(^1\) Weak instruments are variables weakly correlated with the endogenous explanatory variables. Bound, Jaeger, and Baker (1995) pointed out that the weak instrument is not a small-sample by providing an empirical study on weak instruments with 329,000 observations, while Nelson and Startz (1990) and Maddala and Jeong (1992) examined the behavior of the two-stage least squares (TSLS) estimator and showed that the normal approximation of sampling distributions of TSLS estimator can not be good. These findings led many researchers to look for nonstandard approximations to sampling distributions.

Staiger and Stock (1997) was the first paper to consider a classical simultaneous equations model by proposing the so-called local-to-zero parameterization of the coefficients of the instruments in the reduced form equation. Also, Staiger and Stock (1997) showed that, under this local-to-zero framework with the number of instruments fixed, the TSLS and limited information maximum likelihood estimators are inconsistent but converge to nonstandard distributions. Hahn and Kuersteiner (2002) considered the same type model as Staiger and Stock (1997), but generalized Staiger and Stock’s (1997) specification by varying degrees of weakness. Indeed, Hahn and Kuersteiner (2002) considered three cases: (i) the weak case defined by Staiger and Stock (1997), (ii) the nearly weak case, in which the instruments are stronger than the weak case considered by Staiger and Stock (1997), and (iii) the nearly non-identified case, in which the instruments are weaker than the weak case considered by Staiger and Stock (1997). Also, Hahn and Kuersteiner (2002) showed that, for the nearly non-identified case and Staiger and Stock’s (1997) weak case, the TSLS estimators are inconsistent although their limiting distributions exist but not normal, while for the nearly weak, the TSLS estimator is consistent and its limiting distribution is normal. As pointed out by Hahn and Kuersteiner (2002), for the nearly weak case, the limiting distribution does not reflect the type of finite sample moments usually associated with the TSLS estimator,

---

while it was shown by Chao and Swanson (2007) that the weak instrument limit of Staiger and Stock (1997) preserves the exact finite sample moments of TSLS under some regularity conditions. Finally, Cai and Li (2007) extended the work by Hahn and Kuersteiner (2002) for cross-sectional data to panel data model.

There exists a rich literature on linear and nonlinear parametric IV models. It is well known, however, that parametric IV models may be misspecified, and estimators obtained from misspecified models are often inconsistent, so that statistical inferences might lead to inaccurate or wrong conclusions. To overcome this problem, some semi/nonparametric IV models have been proposed over the last several years.\(^2\) For example, Newey and Powell (2003) and Newey, Powell and Vella (1999) considered a nonparametric model with the endogenous variable appearing inside an unknown function and both papers used the series method to estimate the regression function. Cai, Das, Xiong and Wu (2006) aimed on estimating a varying coefficient IV model by allowing for endogenous variables to enter the parametric part of the model, including the nonparametric IV model for discrete endogenous variables tackled by Das (2005), and they proposed a two-stage estimation method to first estimate coefficient functions. Cai and Li (2005) studied the varying coefficient IV model for dynamic panel data and they proposed a so-called nonparametric generalized method of moment to estimate the coefficient functions. Recently, Ai and Chen (2003) focused on an efficient estimation of the parametric components in a general class of semiparametric IV models where the endogenous variable is allowed to appear inside an unknown function, and they mainly considered the \(\sqrt{n}\)-asymptotic normality result for the finite dimensional parameters but they did not provide asymptotic distribution of the nonparametric components. To make models more flexible and to obtain the asymptotic normality of the estimators, Cai and Xiong (2006) studied a class of semiparametric instrumental variables models with the structural function represented by a partially varying coefficient functional form and they proposed efficient (three-step) procedures to estimate both the constant and functional coefficients. Also, Cai and Xiong (2006) established the asymptotic properties of estimators for both the constant and functional coefficients, including consistency and asymptotic normality. However, the aforementioned papers focused on semi/nonparametric instrumental variables models but not for weak instrumentals. Therefore, to our knowledge, there is

\(^2\)Recent work includes, for example, Ai and Chen (2003), Blundell and Powell (2003), Cai, Das, Xiong and Wu (2006), Cai and Li (2005), Cai and Xiong (2006), Das (2005), Newey and Powell (2003), and Newey, Powell and Vella (1999).
no any attempt to consider a semi/nonparametric instrumental variables model for weak instruments.

Functional coefficient models are well known in the statistics and econometrics literature\(^3\) due to their flexibility. Their structure is analogous to that of random coefficients models with an ability of capturing partially heteroskedasticity; see, e.g., Hsiao (2003) and Granger and Teräsvirta (1993). Further, it is worth to pointing out that a functional coefficient model can be regarded as an approximation of the general nonparametric model by a Taylor expansion. A functional coefficient representation for the structural model linearizes the nonparametric function in the endogenous components, yielding a model in which the endogenous components have coefficients depending on unknown functions of predetermined exogenous variables. In such a way, they have appreciable flexibility relative to partially linear models, albeit less general than fully nonparametric models. Since the functional coefficients only depend on exogenous variables, the so called *ill-posed inverse* problem does not exhibit under this setting; see Powell and Newey (2003). Recently, functional coefficient models have been successfully applied to empirical studies in economics. For example, to name just a few, Hong and Lee (2003) explored the inference and forecasting of exchange rates, Juhl (2005) studied the unit root behavior of nonlinear time series models, Li, Huang, Li and Fu (2002) modelled the production frontier using China’s manufactory industry data, and Cai et al. (2006) considered the nonparametric two-stage instrumental variable estimators for returns to education.

The goal of this paper is to consider a simultaneous equations model under a functional coefficient representation for the structural equation of interest with weak instruments and to adopt a local-to-zero assumption as in Hahn and Kuersteiner (2002) on the coefficients of the instruments in the reduced form equation as follows

\[
y_i = g_0(z_{i1}) + \sum_{j=1}^{p} g_j(z_{i1}) x_{ij} + u_i, \quad 1 \leq i \leq n, \tag{1}
\]

\[
x_i = n^{-\alpha} C' z_i + v_i, \quad 1 \leq i \leq n, \tag{2}
\]

where \(\{g_j(\cdot)\}_{j=0}^{p}\) are unspecified smooth coefficient functions, \(\{x_{i\ell}\}_{\ell=1}^{p}\) are endogenous variables, \(z_{i1}\) is the vector of exogenous variables, \(z_i' = (z_{i1}', z_{i2}')\) with \(z_{i2}\) being the vector of

\(^3\)See, for example, the papers by Hastie and Tibshirani (1993), Cai, Fan and Li (2000), Cai (2002), Li, Huang, Li and Fu (2002), Hong and Lee (2003), Juhl (2005), Cai and Li (2005), and Cai et al. (2006) and the references therein.
instrumental variables, $C$ is the parameter matrix, and $0 < \alpha < 1$ controls the degree of weakness.

As mentioned earlier, under this functional coefficient representation, models are linear in the endogenous components with coefficients given by unknown functions of the predetermined variables. Under such a setting, the ill-posed inverse problem disappears. To estimate the coefficient functions $\{g_j(\cdot)\}$, we propose a two-stage estimation procedure similar to that in Cai et al. (2006), described as follows. The first step is to estimate a matrix of unknown parameters of the reduced form equation by using the least squares estimator, and the second step is local linear regression using the estimated reduced forms as regressors. We investigate how the limiting distribution of the resulting estimator changes as the parameterization varies to allow for the different degrees of weakness. The consistency (with the conventional rate of convergence at $\sqrt{nh}$) and the asymptotic normality of the estimator of the coefficient function $g_0(\cdot)$ are established when the instrumental variables are weak for all three cases as in Hahn and Kuersteiner (2002): the weak case considered by Staiger and Stock (1997) ($\alpha = 1/2$), the nearly weak case ($\alpha < 1/2$), and the nearly non-identified case ($\alpha > 1/2$). The consistency (with convergence rate at $n^{1/2-\alpha}h^{1/2}$) and asymptotic normality of the estimator of coefficient functions of endogenous $g_j(\cdot)$ ($j \geq 1$) variables are given when the instrumental variables are the nearly weak case ($\alpha < 1/2$). More importantly, it is shown that the estimators of coefficient functions $g_j(\cdot)$ ($j \geq 1$) of endogenous variables are divergent in the sense that the limiting distribution does not exist, when the instrumental variables are weak as the case considered by Staiger and Stock (1997) ($\alpha = 1/2$), or the nearly non-identified case ($\alpha > 1/2$). By contrast, this differs totally from that for parametric models studied in Hahn and Kuersteiner (2002). These interesting findings seem to be novel in the literature.

The rest of the paper is organized as follows. In Section 2, we introduce the model and propose the nonparametric estimators as well as discuss their large sample results, including the divergence and convergence. For the convergent results, the consistency and asymptotic normality of the estimators are presented in the same section. In Section 3, we examine the finite sample properties of the nonparametric instrumental variables estimator by Monte Carlo simulations. Section 4 provides some preliminary results stated as lemmas and the detailed derivations of main result and its corollaries. Appendix contains the detailed proofs of certain lemmas needed in the proofs of the theorem in Section 2.
2 Statistical Models and Properties

2.1 Setup

We consider the model given in (1) and (2), re-expressed as

\[ y_i = g_0(z_{i1}) + g(z_{i1})'x_i + u_i, \quad 1 \leq i \leq n, \]
\[ x_i = n^{-\alpha}C'z_i + v_i, \quad 1 \leq i \leq n, \]

where \( g(z_{i1}) = (g_1(z_{i1}), \ldots, g_p(z_{i1}))' \), the coefficient functions \( \{g_j(\cdot), 0 \leq j \leq p\} \) are unspecified smooth functions in \( \mathcal{R}^k \) (\( k \geq 1 \), \( z_{i1} \in \mathcal{R}^k \)), \( y_i \) is a scalar dependent variable, \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})' \) is a \( p \times 1 \) column vector of endogenous variables, \( z_{i1} = (z_{i1}, z_{i2}, \ldots, z_{ik})' \) is a \( k \times 1 \) column vector of exogenous variables, \( z_{i2} = (z_{i(k+1)}, z_{i(k+2)}, \ldots, z_{i(k+q)})' \) is a \( q \times 1 \) column vector of instrumental variables, \( z_i = (z_{i1}', z_{i2}')' \), \( C \) is a \( (k+q) \times p \) matrix of unknown parameters, \( v_i = (v_{i1}, v_{i2}, \ldots, v_{ip})' \) is a \( p \times 1 \) column vector of measurement errors, and \( \alpha \) is a known parameter, \( 0 < \alpha < 1 \). Here, we assume that \( z_i \) is uncorrelated with \( u_i \) and uncorrelated with \( v_i \) so that \( z_{i1} \) is a vector of exogenous variables and \( z_{i2} \) is a vector of excluding instrumental variables. That is; \( E(u_i|z_i) = 0 \) and \( E(v_i|z_i) = 0 \).

As showed in Cai et al. (2006), a sufficient condition to identify the model given in (1) is that \( q \geq p \), which is assumed throughout the paper. In what follows, we assume that model (1) is identified. To estimate the nonparametric coefficient functions \( \{g_j(\cdot)\} \), we take conditional expectation on (1) with respect to \( z_i \). It is easy to show that

\[ E(y_i | z_i) = g_0(z_{i1}) + E[x_i|z_i]'g(z_{i1}) = g_0(z_{i1}) + n^{-\alpha}z_i'Cg(z_{i1}) \equiv \pi(z_i)'g^*(z_{i1}), \quad (3) \]

where \( \pi(z_i)' = (1, n^{-\alpha}z_i'C) \) and \( g^*(z_{i1})' = (g_0(z_{i1}), g(z_{i1})') \), which implies that \( \{g_j(\cdot)\} \) are functional coefficients of \( \pi(z_i) \), and \( \{g_j(\cdot)\} \) could be estimated by running a nonparametric regression of \( y_i \) versus \( \pi(z_i) \) if \( \pi(z_i) \) were known. However, \( \pi(z_i) \) is unknown in practice. Therefore, estimating \( \{g_j(\cdot)\} \) requires a two-stage method. A preliminary step is estimation of \( \pi(z_i) \) by a regression of \( x_i \) on \( z_i \), while the next step is the estimation of \( \{g_j(\cdot)\} \) by a regression of \( y_i \) on \( z_i \) and the first step estimated values for \( \hat{\pi}(z_i) \) (the estimator of \( \pi(z_i) \)). This method will be described in detail in the next section.

Note that the class of models given in (1) includes some interesting special cases that arise commonly in empirical research. For example, model (1) includes the nonparametric
IV model with binary endogenous variable considered by Das (2005) and a threshold IV model studied by Caner and Hansen (2004) if $g_j(\cdot)$ is a threshold function.

For simplicity of presentation, we provide some additional definitions and notations. If $W$ is a $p \times q$ matrix, $\text{Vec}(W)$ denotes the $pq \times 1$ vector formed by stacking the columns of $W$ under each; that is, if $W = (W_1, W_2, \cdots, W_q)$, where $W_i$ is a $p \times 1$ vector for $i = 1, \cdots, q$, then $\text{Vec}(W) = (W'_1, W'_2, \cdots, W'_q)'$. Also, $\otimes$ denotes the Kronecker product. Further, we use “$\Rightarrow$” to stand for convergence in distribution and “$\rightarrow_p$” to present convergence in probability. For ease of notation, we consider only the case that $k = 1$ in (1). Extension to the case that $k > 1$ involves no fundamentally new ideas. Note that the asymptotic results for univariate case continue to hold for multivariate case ($k > 1$). For $k = 1$, we change notation from $z_{i1}$ to $z_{i1} \in \mathcal{R}$ throughout this paper.

### 2.2 A Two-Stage Estimator

Given observed data $\{(y_i, x_i, z_i)\}$, our suggested estimation procedure is a two-stage approach, described as follows. The first stage involves estimation of $\pi(z_i)$ by using least squares estimation to model (2) and the second stage is to use a local linear regression to model (3) by replacing $\pi(z_i)$ in (3) by the estimated $\hat{\pi}(z_i)$, denoted by $\hat{\pi}(z_i)$.

We begin with the first stage, where we obtain $\hat{C}$, the estimated value for $C$. To this end, (2) is re-expressed in a matrix form as

$$x = n^{-\alpha} z C + \nu, \tag{4}$$

where $x' = (x_1 \ x_2 \ \cdots \ x_n)$, $z' = (z_1 \ z_2 \ \cdots \ z_n)$, and $\nu' = (\nu_1 \ \nu_2 \ \cdots \ \nu_n)$. Then, using the least squares estimation to reduced form equation (4), we have

$$\hat{C} = n^{\alpha} (z'z)^{-1} z' x. \tag{5}$$

Now, we derive the local linear estimator of $\{g_j(.)\}$. For this purpose, we assume throughout this paper that the functions $\{g_j(.)\}$ have a continuous second derivative at any given point $z_1 \in \mathcal{R}$. By the Taylor expansion for $z_{i1}$ in a neighborhood of $z_1$, $g_j(z_{i1})$ can be approximated by a linear function $\theta_{1,j} + (z_{i1} - z_1) \theta_{2,j}$ with $\theta_{1,j} = g_j(z_1)$ and $\theta_{2,j} = g_j'(z_1) = d g_j(z_1)/dz_1$. Denote $\pi_i = \pi(z_i)$, and $\hat{\pi}_i = \hat{\pi}(z_i) = (1, n^{-\alpha} z'_i \hat{C})'$ as well as $\Pi'_i = (\pi'_i \ (z_{i1} - z_1) \pi'_i)$. Then, the conditional mean in model (3) can be approximated by $E(y_i | z_i) \approx \Pi'_i \Theta$, where
\( \Theta = \Theta(z_1) = (\theta'_1 \theta'_2)' \) with \( \theta_1 = (\theta_{1,0} \cdots \theta_{1,p})' \), and \( \theta_2 = (\theta_{2,0} \cdots \theta_{2,p}) \). The local linear estimator \( \hat{\Theta} \) is defined as the minimizer of the sum of weighted least squares

\[
\sum_{i=1}^{n} \left[ y_i - \sum_{j=0}^{p} (\theta_{1,j} + (z_{i1} - z_1)\theta_{2,j}) \hat{\pi}_{i,j} \right]^2 K_h(z_{i1} - z_1) = \sum_{i=1}^{n} \left[ y_i - \hat{\pi}'_i \Theta \right]^2 K_h(z_{i1} - z_1),
\]

where \( \hat{\pi}_{i,j} \) denotes the \( j \)-th element of \( \hat{\pi}_i \), \( K_h(\cdot) = h^{-1}K(\cdot/h) \), \( K(\cdot) \) is a kernel function on \( \mathcal{R} \), and \( h > 0 \) is the bandwidth at the second step, \( h \rightarrow 0 \) and \( nh \rightarrow \infty \). By minimizing (6) with respect to \( a \) and \( b \), we obtain the local linear estimate of \( \theta_{1,j}(z_1) \) and \( \theta_{2,j}(z_1) \). It follows from the least squares theory that \( \hat{\Theta} = (\hat{\Pi} W \hat{\Pi}')^{-1} \hat{\Pi} W Y \), where \( W = \text{diag}\{K_h(z_{i1} - z_1), \cdots, K_h(z_{n1} - z_1)\} \), \( Y = (y_1, y_2, \cdots, y_n)' \) and \( \hat{\Pi} = (\hat{\Pi}_1 \cdots \hat{\Pi}_n) \) is the estimator of \( \Pi = (\Pi_1 \cdots \Pi_n) \). It is easily verified that \( \hat{\Theta} \) can be re-written as \( \hat{\Theta} = H^{-1}S_n^{-1}T_n \), where \( H = H(h) = \text{diag}\{I_{p+1}, h I_{p+1}\} \), \( I_{p+1} \) is the \( (p+1) \times (p+1) \) identity matrix,

\[
S_n = \frac{1}{n} \sum_{i=1}^{n} z'^{\otimes 2}_{i1} \otimes \hat{\pi}'_{i1} K_h(z_{i1} - z_1), \quad \text{and} \quad T_n = \frac{1}{n} \sum_{i=1}^{n} z^*_{i1} \otimes \hat{\pi}_i K_h(z_{i1} - z_1)y_i
\]

with \( z'^{\otimes 2}_{i1} = z'^{\otimes 2}_{i1} \) and \( z^*_{i1} = z^*_{i1}(h, z_1) = (1, (z_{i1} - z_1)/h)' \).

### 2.3 Distribution Theory

#### 2.3.1 Assumptions and Notations

Set \( H_1 = H_1(n) = \text{diag}\{1, n^a I_{p+1}\} \), \( H_2 = H_2(n) = \text{diag}\{1, n^{1/2} I_{p+1}\} \), and \( H_3 = H_3(n) = H_2 \). Let \( f_1(\cdot) \) be the probability density function of \( z_{i1} \). Define \( \mu_2(K) = \int u^2 K(u) du \), \( \nu_0(K) = \int K^2(u) du \), and \( \nu_2(K) = \int u^2 K^2(u) du \). The following conditions are listed for the asymptotic theory.

**Assumptions:**

1. The kernel \( K(\cdot) \) is symmetric and bounded second order kernel function.

2. \( \{z_i\} \) are independent and identically distributed and \( \Sigma_{zz} = E(z_i z'_i) \) exists and is positive definite. Also, the conditional covariance matrix of \( z_i \) given \( z_{i1} = z_1 \), \( M_2(z_1) = E[z_i z'_i|z_{i1} = z_1] \) is positive definite for a given grid point \( z_1 \).

3. The second order derivative functions \( \{g^{(2)}_j(z_1)\} \) are continuous at a given grid point \( z_1 \).
4. \{(u_i, v_i')\} are independent and identically distributed with the mean zero and conditional covariance matrix of \((u_i, v_i')\) given \(z_i\) is \(\Sigma = \Sigma(z_i) = \begin{pmatrix} \sigma_{uu}(z_i) & \Sigma_{uv}(z_i) \\ \Sigma_{vu}(z_i) & \Sigma_{vv}(z_i) \end{pmatrix}\), positive definite for all \(z_i\).

5. \(h \to 0\) and \(nh \to \infty\).

6. The density function \(f_1(\cdot)\) is continuous and \(f_1(z_1) > 0\) at a given grid point \(z_1\).

7. \(E(((z_{ij1} z_{ij2})^2) < \infty\) for all \(1 \leq j_1, j_2 \leq (1 + q)\).

Assumptions 1 - 3 and 5 - 6 are commonly imposed in local polynomial smoothing methods; see Fan and Gijbels (1996). The asymptotic sampling theory for resulting two-stage estimators is established in Theorem 1 and its corollaries for the consistency, inconsistency, divergency, and asymptotic normality.

To give precisely the distributional results, we need some additional notations. Define \(M_1(z_1) = E(z_i | z_{i1} = z_1)\) and \(M(z_1) = E(z_i^{\otimes 2} | z_{i1} = z_1)\), where \(z_i^* = (1 z_i')\). Set \(\Delta_{uu} = E[\sigma_{uu}(z_i) z_i^{\otimes 2} | z_{i1} = z_1], \Delta_{uv} = g(z_1)'e E[\Sigma_{uv}(z_i) z_i^{\otimes 2} | z_{i1} = z_1], \Delta_{vu} = \Delta_{uv}, \Delta_{vv} = E[g(z_1)'\Sigma_{vv}(z_i) g(z_1) z_i^{\otimes 2} | z_{i1} = z_1]\), and \(\Lambda = \text{diag}\{\Lambda_1, \Lambda_3\}\), where \(\Lambda_3 = E[\Sigma_{vv}(z_i) \otimes z_i^{\otimes 2}]\) and

\[
\Lambda_1 = f_1(z_1) \begin{pmatrix} \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{uu} & \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{uv} \\
\text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{vu} & \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{vv} \end{pmatrix}.
\]

Also, we define a dummy variable \(c(\alpha)\) to characterize different degrees of weakness. \(c(\alpha) = 1\) is for the nearly weak case if \(0 < \alpha < 1/2\), \(c(\alpha) = 2\) stands for the weak case, if \(\alpha = 1/2\), and \(c(\alpha) = 3\) represents the nearly non-identified case, if \(\alpha > 1/2\). Moreover, define, for \(1 \leq j \leq 3\), \(S^{(j)}(z_1) = \text{diag}\{1, \mu_2(K)\} \otimes \Omega^{(j)}(z_1)\), where \(\Omega^{(j)}(z_1) = D_j' M(z_1) D_j, D_{c(\alpha)}^{(j)} = \text{diag}\{1, D_{c(\alpha)}\}, D_1 = C, D_2 = C + \Sigma_{zz}^{-1} Z_v, \text{ and } D_3 = \Sigma_{zz}^{-1} Z_v\) with \(Z_v\) being a \((q + 1) \times p\) matrix of random variables and \(\text{vec}(Z_v) \sim N(0, \Lambda_3)\). Finally, define \(\varepsilon_i = y_i - E(y_i | z_i)\). Then, by (3), \(\varepsilon_i = u_i + v_i g(z_{i1})\) and \(\Delta_\varepsilon = \text{Var}(\varepsilon_i z_i' | z_{i1} = z_1) = E[\sigma_\varepsilon^2(z_i) z_i^{\otimes 2} | z_{i1} = z_1] = \Delta_{uu} + 2\Delta_{uv} + \Delta_{vv}\), where \(\sigma_\varepsilon^2(z_i)\) is the conditional variance of \(\varepsilon_i\) given \(z_i\).

### 2.3.2 Asymptotic Properties

We discuss the asymptotic distribution of the estimator \(\hat{\Theta}\), stated in Theorem 1 with its proof given in Section 4. In particular, we discuss the consistency, inconsistency, divergency, and asymptotic normality of the proposed estimator.
Theorem 1: Under Assumptions 1-7, we have

(i) for \( c(\alpha) = 1 \),

\[
\sqrt{n h} (I_2 \otimes H_1^{-1}) \left[ H \left\{ \hat{\Theta} - \Theta \right\} - \frac{h^2}{2} \left( \mu_2(K) \hat{g}^{(2)}(z_1) \right) \right] \\
\Rightarrow \left[ f_1(z_1)S^{(1)}(z_1) \right]^{-1} (I_2 \otimes D_1^*(z_1)') (Z_{ku} + Z_{kv}),
\]

and (ii) for \( c(\alpha) \geq 2 \),

\[
\sqrt{n h} (I_2 \otimes H_2^{-1}) \left[ H \left\{ \hat{\Theta} - \Theta \right\} - \frac{h^2}{2} \left( \mu_2(K) \hat{g}^{(2)}(z_1) \right) \right] \\
\Rightarrow \left[ f_1(z_1)S^{(c(\alpha))}(z_1) \right]^{-1} (I_2 \otimes D_{c(\alpha)}'(z_1)') (Z_{ku} + Z_{kv}),
\]

where \( Z_{ku} \) and \( Z_{kv} \) are \( 2(q+1) \times 1 \) normal random vectors and the joint distribution of \( Z_{ku} \), \( Z_{kv} \), and \( \text{Vec}(Z_v) \) is \( N(0, \Lambda) \).

Note that \( Z_{ku} \) and \( Z_{kv} \) are independent of \( Z_v \). By Theorem 1, we are ready to have the asymptotic distributions of the estimators \( \hat{g}_0(\cdot) \) and \( \hat{g}(\cdot) \), which are provided in Corollary 1.

Corollary 1: Under Assumptions 1-7, then,

(i) for \( c(\alpha) = 1 \),

\[
\sqrt{n h} H_1^{-1} \left( \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g^{(2)}_0(z_1) \right) \\
\Rightarrow f_1^{-1}(z_1) \Omega^{(1)}(z_1)^{-1} D_1' (I_{q+2}, 0) (Z_{ku} + Z_{kv}),
\]

and (ii) for \( c(\alpha) \geq 2 \),

\[
\sqrt{n h} H_2^{-1} \left( \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g^{(2)}_0(z_1) \right) \\
\Rightarrow f_1^{-1}(z_1) \Omega^{(c(\alpha))}(z_1)^{-1} D_{c(\alpha)}'(I_{q+2}, 0) (Z_{ku} + Z_{kv}),
\]

where \( Z_{ku} \) and \( Z_{kv} \) are given in Theorem 1.

Corollary 2: Under Assumptions 1-7,

(i) for \( c(\alpha) = 1 \),

\[
\sqrt{n h} \left[ \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g^{(2)}_0(z_1) \right] \\
\Rightarrow f_1^{-1}(z_1)(\Omega_{c(\alpha)}, -M_1(z_1)'D_1J_1D_1', 0)(Z_{ku} + Z_{kv}),
\]


9
Corollary 3: It follows clearly from Corollary 2 that the two-stage estimator for $g_0(\cdot)$ is always consistent with same convergence rate at $\sqrt{n}h$ for any $\alpha$ although the magnitudes might be different for different values of $\alpha$. However, the two-stage estimator for $g(\cdot)$ is consistent only for the nearly weak case $0 < \alpha < 1/2$ and the rate of convergence is lower than that for $\hat{g}_0(\cdot)$. Further, the estimator is divergent when $\alpha \geq 1/2$ (for both weak and nearly non-identified cases). Moreover, from Corollary 2, it is easy to obtain the asymptotic normality for $\hat{g}_0(z_1)$ and $\hat{g}(\cdot)$ for the case $0 < \alpha < 1/2$, stated in Corollary 3. These findings seem novel in the literature.

**Corollary 3:** Under Assumptions 1-7, for $0 < \alpha < 1/2$, if $nh^5 = O(1)$, then
\[
\sqrt{n}h \left[ \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K)g_0^{(2)}(z_1) \right] \Rightarrow N(0, \sigma_{g_0}^2(z_1)),
\]
and if $n^{1-2\alpha}h \rightarrow \infty$ and $n^{1-2\alpha}h^5 = O(1)$, then
\[
n^{1/2-\alpha}h^{1/2} \left[ \hat{g}(z_1) - g(z_1) - \frac{h^2}{2} \mu_2(K)g^{(2)}(z_1) \right] \Rightarrow N(0, \Sigma_g(z_1)),
\]
where $\sigma_{g_0}^2(z_1) = f_1^{-1}(z_1)\nu_0(K)e_1' \left[ \Omega_0^{(1)}(z_1) \right]^{-1} D_1' \Delta_1 D_1^* \left[ \Omega_0^{(1)}(z_1) \right]^{-1} e_1, e_1 = (1, 0, \cdots, 0)'$, $\Sigma_g(z_1) = f_1^{-1}(z_1)\nu_0(K)e_2' \left[ \Omega_0^{(1)}(z_1) \right]^{-1} D_1' \Delta_1 D_1^* \left[ \Omega_0^{(1)}(z_1) \right]^{-1} e_2, and e_2' = (0, I_p)$. 
Remark 2: From Corollary 3, we can easily derive the asymptotic mean squared error (AMSE) [the asymptotic variance plus the square of the asymptotic bias] for the estimator \( \hat{g}_0(z_1) \), which is \( \text{AMSE}_0 = \frac{1}{4} h^4 \mu_2^2(K) \left[ \hat{g}_0^{(2)}(z_1) \right]^2 + \sigma_{g_0}^2(z_1)/(nh) \). By minimizing \( \text{AMSE}_0 \) with respect to \( h \), we obtain the optimal bandwidth for \( \hat{g}_0(z_1) \), which is \( h_{0,\text{opt}} = O(n^{-1/5}) \), so that the optimal \( \text{AMSE}_0 \) has an order \( O(n^{-4/5}) \). Similarly, for \( j \geq 1 \), the optimal bandwidth for \( \hat{g}_j(z_1) \) is \( h_{j,\text{opt}} = O(n^{-(1-2\alpha)/5}) \) and the optimal \( \text{AMSE}_j \) is \( O(n^{-4(1-2\alpha)/5}) \), which is larger than \( \text{AMSE}_0 = O(n^{-4/5}) \). This discussion implies that a single value of \( h \) can not make the estimation of both \( g_0(\cdot) \) and \( g(\cdot) \) optimally. Therefore, to estimate both \( g_0(\cdot) \) and \( g(\cdot) \) optimally, some iterative estimation steps are needed. One possible solution is to use the profile least square approach discussed in Cai (2002) and it is beyond the scope of this paper. Of course, it deserves a further investigation in the further studies.

Finally, we would like to point out from Corollary 3 that the asymptotic variances (sandwich form) consist of three terms: the first term \( \Delta_{uu} \) in the meat part \( \Delta_{\varepsilon} \) addresses the variation of measurement error at the second step, the second term \( \Delta_{uv} \) accounts correctly for the asymptotic covariance between the first and second steps, and the third term \( \Delta_{vv} \) measures the variability of the estimated reduced form. By contrast, the presence of the covariance term under this setting is different from some parametric IV estimators; see, for example, Staiger and Stock (1997), Hahn and Kuersteiner (2002), Cai and Li (2007), Li (2006), and Chao and Swanson (2007).

3 Monte Carlo Simulations

To illustrate our modeling procedure, we consider some Monte Carlo simulations. In our computation, we use the Epanechnikov kernel \( K(u) = 0.75(1 - u^2)I(|u| \leq 1) \) as the kernel function. We evaluate the finite sample performances of our estimator in terms of the mean absolute deviation error (MADE)

\[
\mathcal{E}_j = \frac{1}{n_0} \sum_{j=1}^{n_0} |\hat{g}_j(s_j) - g_j(s_j)|,
\]

where \( s_j, 1 \leq j \leq n_0 \) are the regular grid points.

We consider the following data generating model:

\[
y_i = g_0(z_{i1}) + g_1(z_{i2}) x_i + u_i,
\]
\[
x_i = 2N^{-\alpha} z_{i1} - 3N^{-\alpha} z_{i2} + v_i,
\]
where $g_0(x) = 2 \sin(x)$, $g_1(x) = 3 \exp(-(0.5x - 1)^2)$, the exogenous variable $z_{i1}$ is generated from uniform distribution $(-3, 3)$, and the excluded instrument variable $z_{i2}$ is generated from uniform distribution $(-3, 3)$ independently. Finally, $u_i$ and $v_i$ are generated jointly from a standard bivariate normal with the correlation coefficient 0.8. Clearly, $\{(u_i, v_i)\}$ is independent of $z_{i1}$ and $z_{i2}$. But $x_i$ is correlated with $u_i$, since $u_i$ and $v_i$ are correlated. For different degrees of weakness, we consider three cases: $\alpha = 0.2$, 0.5, and 0.7, corresponding to the nearly weak, weak and nearly non-identified cases, respectively. For each case, we consider three sample sizes: $n = 100$, $n = 250$, and $n = 500$ and 500 replications are performed for each sample size.

**Case I: nearly weak** ($\alpha = 0.2$). The results are summarized in Figure 1. For each sample size, the boxplots of the 500 MADE values are plotted in Figures 1(c) for $\hat{g}_0(\cdot)$ and 1(d) for $\hat{g}_1(\cdot)$, respectively. We observe from Figures 1(c) and 1(d) that as the sample size increases,
Figure 2: Simulation results for Case II. (a) The true coefficient function $g_0(\cdot)$ (solid line) and its two-stage local linear estimator (dashed line). (b) Boxplots of the 500 MADE values of $\hat{g}_0(\cdot)$. (c) Boxplots of the 500 MADE values of $\hat{g}_1(\cdot)$.

The MADE value decreases to zero. This implies that both estimators are consistent. Also, we can see that the MADE value for $\hat{g}_0(\cdot)$ decays faster than that for $\hat{g}_1(\cdot)$. These are in line with the asymptotic theory for the proposed estimators that the estimators are consistent and the rate of convergence for $\hat{g}_0(\cdot)$ is faster than that for $\hat{g}_1(\cdot)$. Figures 1(a) and 1(b), respectively display the true coefficient functions $g_0(\cdot)$ and $g_1(\cdot)$ (solid line) with their two-stage local linear estimators (dashed line) for $n = 500$ from a typical sample. The typical sample is selected in such a way that its total MADE value ($= \mathcal{E}_0 + \mathcal{E}_1$) equals to the median of the 500 replications. Overall, the proposed modeling procedure performs fairly well.

**Case II: weak ($\alpha = 0.5$).** The settings are same as those for Case I. The results are reported in Figure 2. For each sample size, the boxplots of the 500 MADE values are respectively plotted in Figure 2(b) for $\hat{g}_0(\cdot)$ and Figure 2(c) for $\hat{g}_1(\cdot)$. We observe from Figure 2(b) that as the sample size increases, the MADE for $\hat{g}_0(\cdot)$ value becomes smaller. This concludes
that \( \tilde{g}_0(\cdot) \) is consistent. But the MADE for \( \tilde{g}_1(\cdot) \) in Figure 2(c) has an increasing trend as \( n \) becomes larger, which implies that the estimator for \( g_1(\cdot) \) is divergent. Figure 2(a) displays the true coefficient function \( g_0(\cdot) \) (solid line) and its the two-stage local linear estimator (dashed line) for \( n = 500 \) from a typical sample. The typical sample is selected in such a way that its MADE value (\( = E_0 \)) equals to the median the 500 replications.

Case III: nearly non-identified (\( \alpha = 0.7 \)). The settings are same as those for Case II. The results are presented in Figure 3. The same conclusion as that for Case II can be made. Further, we can observe from Figure 3(c) that the divergent rate is slightly faster than that for the weak case (\( \alpha = 0.5 \)) due to the weaker correlation between the endogenous variable and the instruments.
4 Derivations

This section is devoted to the proofs of Theorem 1 and Corollaries 1 - 3. To prove Theorem 1, we first consider the asymptotic behavior of \( \hat{S}_n \) in (7). The result is stated in the following lemma, which will be used subsequently. The proofs of this lemma and other lemmas are given in the appendix.

**Lemma 1:** Under Assumptions 1-7, then

\[
(I_2 \otimes H_{c(\alpha)}) \hat{S}_n (I_2 \otimes H_{c(\alpha)}) \Rightarrow f_1(z_1) S^{(c(\alpha))}(z_1),
\]

where \( S^{(j)}(z_1) \) is Section 2.

Before we embrace on establishing the asymptotic properties of the resulting estimator, first, we decompose \( H[\hat{\Theta} - \Theta] \) into three terms as

\[
H[\hat{\Theta} - \Theta] \equiv \hat{S}_n^{-1} [P_n + Q_n + R_n],
\]

where with \( G = (\pi_{1}^t g^* (z_{11}), \ldots, \pi_{n}^t g^* (z_{n1}))' \), \( P_n = n^{-1}H^{-1}\hat{\Pi}W(Y - G) \), \( Q_n = n^{-1}H^{-1}\hat{\Pi}W(G - \Pi'\Theta) \), and \( R_n = n^{-1}H^{-1}\hat{\Pi}W(\Pi'\Theta - \hat{\Pi}'\Theta) \). The reason of doing the above decomposition is to show that \( P_n \) contributes to only the asymptotic variance, \( Q_n \) is the resource of the bias, and \( R_n \) is negligible comparing with \( P_n \), which are presented in the following lemmas.

**Lemma 2:** Under Assumptions 1-6, we have

\[
\sqrt{n} h (I_2 \otimes H_{c(\alpha)}) P_n \Rightarrow (I_2 \otimes D_{c(\alpha)}^*) (Z_{ku} + Z_{kg}),
\]

where the distributions of \( Z_{ku} \) and \( Z_{kg} \) are given in Theorem 1.

**Lemma 3:** Under Assumptions 1-7, we have

(i) for \( c(\alpha) \leq 2 \),

\[
(I_2 \otimes H_{c(\alpha)}) Q_n = f_1(z_1) \left( \mu_2(K) \right) \otimes (D_{c(\alpha)}^* M(z_1) D_{c(\alpha)}^* H_{c(\alpha)}^{-1} g^{(2)}(z_1) \left[ \frac{h^2}{2} + o_p(h^2) \right]),
\]

and (ii) for \( c(\alpha) = 3 \),

\[
(I_2 \otimes H_{3}) Q_n = f_1(z_1) \left( \mu_2(K) \right) \otimes (D_{3}^* M(z_1) D_{1}^* H_{1}^{-1} g^{(2)}(z_1) \left[ \frac{h^2}{2} + o_p(h^2) \right]).
\]
Thus, by Lemmas 1 and 3, under Assumptions 1-7, we have Lemma 4:

\[ n^{1/2} (I_2 \otimes H_{c(\alpha)}) R_{n,1} = O_p(1), \quad \text{and} \quad h^{-1} n^{1/2} (I_2 \otimes H_{c(\alpha)}) R_{n,2} = O_p(1). \]

Next, we proceed with the proof of Theorem 1 and its corollaries.

Proof of Theorem 1: It is easy to conclude from (8) that

\[ H \left[ \hat{\Theta} - \Theta \right] - \hat{S}_n^{-1} Q_n - \hat{S}_n^{-1} R_n = \hat{S}_n^{-1} P_n. \]  \hspace{1cm} (9)

First, we consider the nearly weak case \(0 < \alpha < 1/2\). To this end, by Lemmas 1 and 4,

\[ \sqrt{n} h (I_2 \otimes H_1^{-1}) \hat{S}_n^{-1} R_n = h^{1/2} \left[ (I_2 \otimes H_1) \hat{S}_n (I_2 \otimes H_1) \right]^{-1} n^{1/2} (I_2 \otimes H_1(n)) R_n \rightarrow^p 0, \]

and by Lemmas 1 and 2,

\[ \sqrt{n} h (I_2 \otimes H_1^{-1}) \hat{S}_n^{-1} P_n = \left[ (I_2 \otimes H_1) \hat{S}_n (I_2 \otimes H_1) \right]^{-1} n^{1/2} h^{1/2} (I_2 \otimes H_1) P_n \]

\[ \Rightarrow \left[ f_1(z_1) S^{(1)}(z_1) \right]^{-1} \left( I_2 \otimes D_1' \right) (Z_{ku} + Z_{kv}). \]

By Lemmas 1 and 3,

\[ (I_2 \otimes H_1^{-1}(n)) \hat{S}_n^{-1} Q_n \]

\[ = \left[ (I_2 \otimes H_1) \hat{S}_n (I_2 \otimes H_1) \right]^{-1} (I_2 \otimes H_1(n)) Q_n \]

\[ = \frac{h^2}{2} \left[ f_1(z_1) S^{(1)}(z_1) \right]^{-1} f_1(z_1) \left( \mu_2(K) \right) \otimes \Omega^{(1)}(z_1) H_1^{-1} g^{(2)}(z_1) + o_p(h^2) \]

\[ = \frac{h^2}{2} \left( \mu_2(K) g^{(2)}(z_1) \right) + o_p(h^2). \]

Thus,

\[ \sqrt{n} h (I_2 \otimes H_1^{-1}) \left[ H \left( \hat{\Theta} - \Theta \right) - \frac{h^2}{2} \left( \mu_2(K) g^{(2)}(z_1) \right) \right] \]

\[ \Rightarrow \left[ f_1(z_1) S^{(1)}(z_1) \right]^{-1} \left( I_2 \otimes D_1' \right) (Z_{ku} + Z_{kv}). \]
Second, we consider the weak case ($\alpha = 1/2$). Similar to the above arguments, we have

$$\sqrt{n} h(I_2 \otimes H^{-1}_2)^{-1}S_n^{-1}R_n \to^p 0,$$

$$\sqrt{n} h(I_2 \otimes H^{-1}_2)^{-1}S_n^{-1}P_n \to [f_1(z_1)S^{(2)}(z_1)]^{-1} (I_2 \otimes D^*_2) (Z_{ku} + Z_{kvg}),$$

and

$$(I_2 \otimes H^{-1}_2)^{-1}Q_n = \frac{h^2}{2} \left[ S^{(2)}(z_1) \right]^{-1} \left( \begin{pmatrix} \mu_2(K) \\ 0 \end{pmatrix} \right) \otimes (D^*_2 M(z_1) D^*_1) H^{-1}_2 g^{(2)}(z_1) + o_p(h^2).$$

Then, by (9),

$$\sqrt{n} h(I_2 \otimes H^{-1}_2)H \left[ \Theta - \Theta \right] - \text{Bias} \to [f_1(z_1)S^{(2)}(z_1)]^{-1} (I_2 \otimes D^*_2) (Z_{ku} + Z_{kvg}),$$

where

$$\text{Bias} = \frac{1}{2} n^{1/2} h^{5/2} (S^{(2)}(z_1))^{-1} \left( \begin{pmatrix} \mu_2(K) \\ 0 \end{pmatrix} \right) \otimes (D^*_2 M(z_1) D^*_1 H^{-1}_2 g^{(2)}(z_1)).$$

Now, we calculate the bias term $\text{Bias}$. Indeed,

$$\text{Bias} = \frac{1}{2} n^{1/2} h^{5/2} \left( \begin{pmatrix} \mu_2(K) \\ 0 \end{pmatrix} \right) \otimes \left[ \Omega^{(2)}(z_1)^{-1} D^*_2 M(z_1) \left( \begin{pmatrix} g^{(2)}_0 \\ 0 \end{pmatrix} \right) + o_p(1) \right]$$

$$= \frac{1}{2} n^{1/2} h^{5/2} \left( \begin{pmatrix} \mu_2(K) \\ 0 \end{pmatrix} \right) \otimes \left[ \left( \begin{pmatrix} g^{(2)}_0 \\ 0 \end{pmatrix} \right) + o_p(1) \right]$$

$$= \frac{1}{2} n^{1/2} h^{5/2} \left[ \left( \begin{pmatrix} \mu_2(K) g^{(2)}_0(z_1) \\ 0 \end{pmatrix} \right) + o_p(1) \right].$$

Therefore,

$$\sqrt{n} h(I_2 \otimes H^{-1}_2) \left[ H \left\{ \Theta - \Theta \right\} - \frac{h^2}{2} \left( \begin{pmatrix} \mu_2(K) g^{(2)}_0(z_1) \\ 0 \end{pmatrix} \right) \right]$$

$$\Rightarrow [f_1(z_1)S^{(2)}(z_1)]^{-1} (I_2 \otimes D^*_2) (Z_{ku} + Z_{kvg}),$$

Similar to the proof for the case $\alpha = 1/2$, we can establish the case $1/2 < \alpha$. Hence, the proof of Theorem 1 is complete.

**Proof of Corollary 1:** From Theorem 1, it suffices to compute each component of the limiting distribution given in Theorem 1. To this end, some simple algebras lead to

$$[f_1(z_1)S^{(c(\alpha))}(z_1)]^{-1} (I_2 \otimes D^*_c) (Z_{ku} + Z_{kvg})$$

$$= \left[ f_1^{-1}(z_1) \begin{pmatrix} 0 & 0 \\ 1 & \mu_2(K) \end{pmatrix} \right]^{-1} \otimes \left[ \Omega^{(c(\alpha))}(z_1)^{-1} \right] (I_2 \otimes D^*_c) (Z_{ku} + Z_{kvg}).$$
By the inverse of a partitioned matrix,

\[ f_1^{-1}(z_1) \Omega^{(c(\alpha))}(z_1)^{-1} D_{c(\alpha)}^\prime \left( I_{q+2} \ 0 \right) (Z_{ku} + Z_{kvg}) \]

By Corollary 1, we have proved Corollary 2.

**Proof of Corollary 2:** By the inverse of a partitioned matrix,

\[ \Omega^{(c(\alpha))}(z_1)^{-1} = \begin{pmatrix} \Omega_{c(\alpha)} & -M_1(z_1)' D_{c(\alpha)} J_{c(\alpha)} \\ -J_{c(\alpha)} D_{c(\alpha)}' M_1(z_1) & J_{c(\alpha)} \end{pmatrix}, \]

which implies that

\[ f_1^{-1}(z_1) \Omega^{(c(\alpha))}(z_1)^{-1} D_{c(\alpha)}^\prime \left( I_{q+2} \ 0 \right) (Z_{ku} + Z_{kvg}) \]

\[ = f_1^{-1}(z_1) \Omega^{(c(\alpha))}(z_1)^{-1} D_{c(\alpha)}^\prime \left( \begin{pmatrix} 1 & 0 & 0 \\ 0 & I_{q+1} & 0 \end{pmatrix} (Z_{ku} + Z_{kvg}) \right) \]

\[ = \begin{pmatrix} f_1^{-1}(z_1) \left( \Omega_{c(\alpha)} - M_1(z_1)' D_{c(\alpha)} J_{c(\alpha)} D_{c(\alpha)}^\prime 0 \right) (Z_{ku} + Z_{kvg}) \\ f_1^{-1}(z_1) \left( -J_{c(\alpha)} D_{c(\alpha)}^\prime M_1(z_1) J_{c(\alpha)} D_{c(\alpha)}^\prime 0 \right) (Z_{ku} + Z_{kvg}) \end{pmatrix}. \]

By Corollary 1, we have proved Corollary 2.

**Proof of Corollary 3:** It is easy to see from Lemma A.1 in Appendix that \( \left( \begin{pmatrix} Z_{ku} \\ Z_{kvg} \end{pmatrix} \right) \sim N(0, \Lambda_1) \), and \( \left( I_{q+2} \ 0 \right) (Z_{ku} + Z_{kvg}) \sim N(0, \Lambda_4) \), where \( \Lambda_4 = f_1(z_1) \nu_0(K) \Delta_c \). By the fact that

\[ \Omega^{(1)}(z_1)^{-1} = \begin{pmatrix} \Omega_1 & -M_1(z_1)' C J_1 \\ -J_1 C' M_1(z_1) & J_1 \end{pmatrix}, \]

we have

\[ f_1^{-2}(z_1) \Omega^{(1)}(z_1)^{-1} D_{c(\alpha)}^\prime \Lambda_3 D_{c(\alpha)}^\prime \Omega^{(1)}(z_1)^{-1} = \begin{pmatrix} \sigma_{g0}^2(z_1) & \Sigma_{g0,g}(z_1) \\ \Sigma_{g0,g}(z_1) & \Sigma_g(z_1) \end{pmatrix} \]

for some \( \Sigma_{g0,g}(z_1) \). Thus, Corollary 3 holds from Corollary 2. Hence, Corollary 3 is proved.
5 Conclusions

This paper considers a nonparametric structural model that satisfies a functional coefficient representation under the weak instrumental assumptions as Staiger and Stock (1997) and Hahn and Kuersteiner (2002) by allowing the different degrees of weakness. This model representation can be regarded as a generalization of classical random coefficients models and is useful in applications. In particular, under this representation the model overcomes the so-called ill-posed problem of other structural models while retaining appreciable flexibility over partially linear models. A two-step local linear estimator is developed to estimate the coefficient functions. Asymptotic properties including consistency and asymptotic normality and divergency are derived. Finally, some future researches related to this work include deriving asymptotic properties for the linear component of a partially linear case of the model, choosing optimal weak instruments, considering the case when the number of weak instruments goes to infinity, and selecting the optimal bandwidth, as well as obtaining the optimality advocated in Remark 2.

Appendix

Throughout this appendix, we use the same notations as introduced in Sections 2 and 4. Before we embrace on the proofs of Lemmas 1-4, we first establish three preliminary results below. Also, we employ the following notations. Define \( C_\beta = \text{diag}\{1, n^\beta C\} \) and \( \hat{C}_\beta = \text{diag}\{1, n^\beta \hat{C}\} \).

Lemma A.1: Let \( \eta_i' = z_i^* \otimes z_i^* \). Then, under Assumptions 1-4 and 6, we have

\[
n^{-1/2} \sum_{i=1}^{n} (h^{1/2} \eta_i' u_i K_h(z_{i1} - z_1), h^{1/2} \eta_i' v'_i g(z_{i1}) K_h(z_{i1} - z_1), z_i v'_i) \Rightarrow (Z_{ku}, Z_{kvg}, Z_v),
\]

where the joint distribution of \( Z_{ku}, Z_{kvg}, \) and \( \text{Vec}(Z_v) \) is \( N(0, \Lambda) \).

Proof: It is clear that to establish the lemma, it suffices to show that \( n^{-1/2} \sum_{i=1}^{n} \xi_i \Rightarrow N(0, \Lambda) \), where

\[
\xi_i = \begin{pmatrix}
    h^{1/2} \eta_i' u_i K_h(z_{i1} - z_1) \\
    h^{1/2} \eta_i' v'_i g(z_{i1}) K_h(z_{i1} - z_1) \\
    \text{Vec}(z_i v'_i)
\end{pmatrix} \equiv \begin{pmatrix}
    \xi_{i1} \\
    \xi_{i2} \\
    \xi_{i3}
\end{pmatrix}.
\]

Clearly, \( E[\xi_i] = 0 \). Since \( \{(z_i, u_i, v_i)\} \) are independent and identically distributed, then, so are \( \{\xi_i\} \). It follows from the central limit theorem and the kernel smoothing technique (e.g.,
Similarly, one can show easily that
\[ \text{Cov}(\xi_i) = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}, \]
where \( A_{11} = \text{Var}(\xi_{i1}), \) \( A_{12} = \text{Cov}(\xi_{i1}, \xi_{i2}), \) \( A_{13} = \text{Cov}(\xi_{i1}, \xi_{i3}), \) \( A_{22} = \text{Var}(\xi_{i2}), \) \( A_{23} = \text{Cov}(\xi_{i2}, \xi_{i3}), \) \( A_{33} = \text{Var}(\xi_{i3}), \) \( A_{21} = A'_{12}, \) \( A_{31} = A'_{13}, \) and \( A_{32} = A'_{23}. \) By Assumptions 1-4 and 6, one can show easily that \( \eta'_i \eta_i = z^{*\otimes2}_i \otimes z^{*\otimes2}_i, \) \( A_{33} = E[\Sigma_{uv}(z_i) \otimes z_i z'_i] = \Lambda_3, \)
\[
A_{11} = h \sum_{i=1}^{n} E \left[ z^{*\otimes2}_i \otimes z^{*\otimes2}_i \right] K^2_h(z_{i1} - z_1) = f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{uu} + o(1),
\]
\[
A_{12} = h \sum_{i=1}^{n} E \left[ g'(z_{i1})v_i u_i z^{*\otimes2}_i \otimes z^{*\otimes2}_i K^2_h(z_{i1} - z_1) \right] = f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{uv} + o(1),
\]
and
\[
A_{22} = h \sum_{i=1}^{n} E \left[ g(z_{i1})v_i v'_i g'(z_{i1})z^{*\otimes2}_i \otimes z^{*\otimes2}_i K^2_h(z_{i1} - z_1) \right] = f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \Delta_{vv} + o(1).
\]
Similarly, one can show easily that
\[
A_{13} = h^{1/2} \sum_{i=1}^{n} E \left[ \left( \frac{u_i v_i}{h} \right) \otimes \left( \frac{z'_i}{z_i} \right) K_h(z_{i1} - z_1) \right] = h^{1/2} f_1(z_1) E \left[ \left( \Sigma_{uv}(z_i) \right) \otimes \left( \frac{z'_i}{z_i} \right) \left| z_{i1} = z_1 \right\} + o(h^{1/2}) = o(1), \right.
\]
and
\[
A_{23} = h^{1/2} \sum_{i=1}^{n} E \left[ \left( \frac{g'(z_{i1}) v_i v'_i}{h} \right) \otimes \left( \frac{z'_i}{z_i} \right) K_h(z_{i1} - z_1) \right] = h^{1/2} f_1(z_1) E \left[ \left( \frac{g'(z_{i1}) \Sigma_{uv}(z_i)}{h} \right) \otimes \left( \frac{z'_i}{z_i} \right) \left| z_{i1} = z_1 \right\} + o(h^{1/2}) = o(1). \right.
\]
Therefore, we prove the lemma.

**Lemma A.2:** Under Assumptions 1-4 and 6,
\[ n^{1/2-a} \left[ \hat{C} - C \right] \Rightarrow \Sigma_{zz}^{-1} Z_v. \]
Moreover, for \( c(\alpha) \leq 2, \) \( \hat{C} \Rightarrow D_{c(\alpha)}, \) and for \( c(\alpha) = 3, \) \( n^{1/2-a} \hat{C} \Rightarrow D_{c(\alpha)}. \)

**Proof:** Since \( \{z_i\} \) are iid, it follows by a law of large numbers that \( n^{-1} z'z \rightarrow^p \Sigma_{zz}. \) It follows from equations (4) and (5) that \( \hat{C} = C + n^{a-1/2} (n^{-1} z'z)^{-1} (n^{-1/2} z'v), \) which, in conjunction

20
with Lemma A.1, implies that \( n^{1/2-\alpha}(\hat{C} - C) \Rightarrow \Sigma_{zz}^{-1}Z_v \). In particular, for \( 0 < \alpha < 1/2, \) \( \hat{C} \rightarrow^p C \). Therefore, Lemma A.2 holds.

**Lemma A.3:** Under Assumptions 1-7, we have

\[
B_{i,j} = n^{-1} \sum_{i=1}^{n} z_i^{\otimes 2} [z_{i1}]^j K_h(z_{i1} - z_1) \quad \rightarrow^p \mu_j(K)f_1(z_1)M_l(z_1),
\]

where \( z_i^{\otimes 2} = z_i z'_i \) and \( z_i^{\otimes 1} = z_i \).

**Proof:** It follows from the kernel smoothing technique by computing the mean and variance; see Fan and Gijbels (1996).

**Proof of Lemma 1:** First, we consider the case when \( c(\alpha) \leq 2 \). By (7), we can re-write \( \hat{S}_n \) as follows

\[
\hat{S}_n = (I_2 \otimes H_{c(\alpha)}^{-1}) \left( I_2 \otimes \hat{C}_0 \right) \left( \begin{array}{cc} B_0 & B_1 \\ B_1 & B_2 \end{array} \right) \left( I_2 \otimes \hat{C}_0 \right) (I_2 \otimes H_{c(\alpha)}^{-1}),
\]

where \( B_0 = \begin{pmatrix} B_{0,0} & B_{1,0} \\ B_{1,0} & B_{2,0} \end{pmatrix} \), \( B_1 = \begin{pmatrix} B_{0,1} & B_{1,1} \\ B_{1,1} & B_{2,1} \end{pmatrix} \), and \( B_2 = \begin{pmatrix} B_{0,2} & B_{1,2} \\ B_{1,2} & B_{2,2} \end{pmatrix} \). Then, it follows from Lemmas A.2 and A.3 that \( (I_2 \otimes H_{c(\alpha)})\hat{S}_n(I_2 \otimes H_{c(\alpha)}) \Rightarrow f_1(z_1)S^{(c(\alpha))}(z_1) \). Now, we consider the case when \( c(\alpha) = 3 \). Since

\[
(I_2 \otimes H_3)\hat{S}_n(I_2 \otimes H_3) = (I_2 \otimes \hat{C}_1^{-1/2-a}) \left( \begin{array}{cc} B_0 & B_1 \\ B_1 & B_2 \end{array} \right) \left( I_2 \otimes \hat{C}_1^{-1/2-a} \right),
\]

then from Lemmas A.2 and A.3, we have

\[
(I_2 \otimes H_3)\hat{S}_n(I_2 \otimes H_3) \Rightarrow f_1(z_1)S^{(3)}(z_1).
\]

The proof of Lemma 1 is complete.

**Proof of Lemma 2:** It is easy to see that

\[
P_n = n^{-1} \sum_{i=1}^{n} z_i^{\otimes 1} \otimes \pi_i K_h(z_{i1} - z_1)(u_i + v'_i g(z_{i1})).
\]

First, for the case when \( c(\alpha) \leq 2 \), one has

\[
\sqrt{n h} P_n = (I_2 \otimes H_{c(\alpha)}^{-1}) \left( I_2 \otimes \hat{C}_0 \right) n^{-1/2} h^{1/2} \sum_{i=1}^{n} \eta_i K_h(z_{i1} - z_1)(u_i + v'_i g(z_{i1})),
\]

where \( \eta_i \) is defined in Lemma A.1. Then, from Lemmas A.1 and A.2, we obtain

\[
\sqrt{n h} (I_2 \otimes H_{c(\alpha)})P_n \Rightarrow (I_2 \otimes D_{c(\alpha)}')(Z_ku + Z_kv),
\]

21
Next, for the case when \( c(\alpha) = 3 \), we re-express \( \sqrt{n} h P_n \) as

\[
\sqrt{n} h P_n = (I_2 \otimes H_3^{-1}) \left( I_2 \otimes \tilde{C}_{1/2-\alpha}' \right) n^{-1/2} h^{1/2} \sum_{i=1}^{n} \eta_i K_h(z_{i1} - z_1)(u_i + v_i' g(z_{i1})).
\]

Then, it follows from Lemmas A.1 and A.2 that

\[
\sqrt{n} h (I_2 \otimes H_3) P_n \Rightarrow (I_2 \otimes D_3^*) (Z_{ku} + Z_{kvg}).
\]

This completes the proof of Lemma 2.

**Proof of Lemma 3:** For \( z_{i1} \) in a neighborhood of \( z_1 \), by the Taylor expansion,

\[
g_j(z_{i1}) = g_j(z_1) + (z_{i1} - z_1) g_j^{(1)}(z_1) + \frac{1}{2} (z_{i1} - z_1)^2 g_j^{(2)}(z_1) + o_p(h^2).
\]

Then,

\[
Q_n = n^{-1} \sum_{i=1}^{n} z_{i1} \otimes \tilde{\pi}_i K_h(z_{i1} - z_1) \pi_i' \left[ \frac{1}{2} (z_{i1} - z_1)^2 g^{*\prime}(z_1) + o_p(h^2) \right].
\]

For the case that \( c(\alpha) \leq 2 \), one has

\[
(I_2 \otimes H_{c(\alpha)}) Q_n = \frac{1}{2} h^2 \left( I_2 \otimes \tilde{C}_{0}' \right) \left( B_4 \over B_5 \right) \otimes D_1^* H_{c(\alpha)}^{-1} g^{*\prime}(z_1) + o_p(h^2),
\]

where \( B_4 = \left( B_{0,2} \quad B_{1,2}' \quad B_{2,2}' \right) \) and \( B_5 = \left( B_{0,3} \quad B_{1,3}' \quad B_{2,3}' \right) \) with \( B_{i,j} \) defined in Lemma A.3. An application of Lemmas A.2 and A.3 leads to

\[
(I_2 \otimes H_{c(\alpha)}) Q_n = \frac{1}{2} h^2 \left( I_2 \otimes \tilde{C}_{0}' \right) \left( f_1(z_1) \mu_2(K) M(z_1) D_1^* \right) H_{c(\alpha)}^{-1} g^{*\prime}(z_1) + o_p(h^2).
\]

Similarly, for the case that \( c(\alpha) = 3 \),

\[
(I_2 \otimes H_3) Q_n = \frac{1}{2} h^2 \left( I_2 \otimes \tilde{C}_{1/2-\alpha}' \right) \left( B_4 \over B_5 \right) \otimes D_1^* H_1^{-1} g^{*\prime}(z_1) + o_p(h^2)
\]

\[
= \frac{1}{2} h^2 \left( I_2 \otimes D_3^* \right) \left( f_1(z_1) \mu_2(K) M(z_1) D_1^* \right) H_1^{-1} g^{*\prime}(z_1) + o_p(h^2)
\]

\[
= \frac{1}{2} h^2 f_1(z_1) \left( \mu_2(K) \right) \otimes (D_3^* M(z_1) D_1^*) H_1^{-1} g^{*\prime}(z_1) + o_p(h^2).
\]

This proves Lemma 3.

**Proof of Lemma 4:** Similar to Lemma 3, we first consider the case when \( c(\alpha) \leq 2 \). To this end, we rewrite \( R_{n,1} \) as

\[
n^{1/2} (I_2 \otimes H_{c(\alpha)}) R_{n,1} = \left( I_2 \otimes \tilde{C}_{0}' \right) \left( B_4 \over B_5 \right) n^{1/2-\alpha} \left[ C - \tilde{C} \right] g(z_1),
\]

\[22\]
where $B_6 = \begin{pmatrix} B'_{1,0} \\ B_{2,0} \end{pmatrix}$ and $B_7 = \begin{pmatrix} B'_{1,1} \\ B_{2,1} \end{pmatrix}$. Applying Lemmas A.2 and A.3, we have

$$n^{1/2}(I_2 \otimes H_{c(\alpha)}) R_{n,1} \Rightarrow -f_1(z_1) \left(\begin{array}{c} 1 \\ 0 \end{array}\right) \otimes \left[ D'_{c(\alpha)} \begin{pmatrix} M_1(z_1)' \\ M_2(z_1) \end{pmatrix} \Sigma_{zz}^{-1} Z_v g(z_1) \right].$$

By an analogue,

$$h^{-1} n^{1/2}(I_2 \otimes H_{c(\alpha)}) R_{n,2} = (I_2 \otimes \tilde{C}_{1/2-\alpha}) \begin{pmatrix} B'_{6} \\ B_7 \end{pmatrix} n^{1/2-\alpha} (C - \tilde{C}) g'(z_1)
\Rightarrow -f_1(z_1) \left(\begin{array}{c} 1 \\ 0 \end{array}\right) \otimes \left[ D'_{3} \begin{pmatrix} M_1(z_1)' \\ M_2(z_1) \end{pmatrix} \Sigma_{zz}^{-1} Z_v g(z_1) \right],$$

where $B_8 = \begin{pmatrix} B'_{1,2} \\ B_{2,2} \end{pmatrix}$. Next, for the case that $c(\alpha) = 3$, by the same token,

$$n^{1/2}(I_2 \otimes H_3(n)) R_{n,1} = (I_2 \otimes \tilde{C}_{1/2-\alpha}) \begin{pmatrix} B'_{6} \\ B_7 \end{pmatrix} n^{1/2-\alpha} (C - \tilde{C}) g(z_1)
\Rightarrow -f_1(z_1) \left(\begin{array}{c} 1 \\ 0 \end{array}\right) \otimes \left[ D'_{3} \begin{pmatrix} M_1(z_1)' \\ M_2(z_1) \end{pmatrix} \Sigma_{zz}^{-1} Z_v g(z_1) \right].$$

and

$$h^{-1} n^{1/2}(I_2 \otimes H_3(n)) R_{n,2} = (I_2 \otimes \tilde{C}_{1/2-\alpha}) \begin{pmatrix} B'_{6} \\ B_7 \end{pmatrix} n^{1/2-\alpha} (C - \tilde{C}) g'(z_1)
\Rightarrow -f_1(z_1) \left(\begin{array}{c} 0 \\ \mu_2(K) \end{array}\right) \otimes \left[ D'_{3} \begin{pmatrix} M_1(z_1)' \\ M_2(z_1) \end{pmatrix} \Sigma_{zz}^{-1} Z_v g'(z_1) \right].$$

This accomplishes the proof of the lemma.

References


