

Convergency and divergency of functional coefficient weak instrumental variables models*

ZONGWU CAI AND HENONG LI

In this paper, we consider a functional coefficient model under weak instrument assumptions as in Staiger and Stock (1997) and Hahn and Kuersteiner (2002). 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. We investigate how the limiting distribution of the proposed nonparametric estimator changes as the degree of weakness of instruments varies. As a result, our new theoretical findings are that the possible convergency of the proposed nonparametric estimator 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.

KEYWORDS AND PHRASES: Discontinuity, Divergence, Endogeneity, Functional coefficient model, Local linear fitting, Simultaneous equations, Weak instrumental variables.

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. A non-exhaustive list of important recent contributions to this growing literature include Angrist and Krueger (1991), Cai, Fang and Li (2007), Campbell (2003), Chao and Swanson (2007), Hahn, Hausman and Kuersteiner (2004), Hahn and Kuersteiner (2002), Hausman, Stock and Yogo (2005), Li (2006), Mavroeidis (2004), Nason and Smith (2005), Neeley, Roy and White-man (2001), Staiger and Stock (1997), Stock (2002), Stock

and Wright (2000), Stock, Wright and Yogo (2002), Woodford (2003), Yogo (2004), and the references contained therein. 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 problem 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. This idea is similar to the local-to-unity in nonstationary time series. 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, while 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 Chao and Swanson (2007) showed that the weak instrument limit of Staiger and Stock (1997) preserves the exact finite sample moments of TSLS under some regularity conditions. Finally, Cai, Fang and Li (2007) extended the work by Hahn and Kuersteiner (2002) for cross-sectional data to panel data model.

*We thank Qiwei Yao for his constructive comments on an early version of this paper. We also thank an associate editor and two anonymous referees for their helpful comments that greatly improved the paper. 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.

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

$$(1) \quad y_i = g_0(\mathbf{z}_{i1}) + \sum_{j=1}^p g_j(\mathbf{z}_{i1}) x_{ij} + u_i, \quad 1 \leq i \leq n,$$

$$(2) \quad \mathbf{x}_i = n^{-\alpha} \mathbf{C}' \mathbf{z}_i + \mathbf{v}_i, \quad 1 \leq i \leq n,$$

where $\{g_j(\cdot)\}_{j=0}^p$ are unspecified smooth coefficient functions, $\{x_{ij}\}_{j=1}^p$ are endogenous variables ($\{x_{ij}\}_{j=1}^p$ and u_i are correlated), \mathbf{z}_{i1} is the vector of exogenous variables, $\mathbf{z}'_i = (\mathbf{z}'_{i1}, \mathbf{z}'_{i2})$ with \mathbf{z}_{i2} being the vector of instrumental variables, \mathbf{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 driven 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 semiparametric estimation procedure similar to that in Cai, Das, Xiong and Wu (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 estimation, and the second step is local linear regression using the estimated reduced forms as regressors. We investigate how the limiting distribution of the resulting nonparametric estimator changes as the different degree of weakness (α) varies. The consistency (with the conventional rate of convergence at $\sqrt{n\hat{h}}$) and the asymptotic normality of the estimator of the intercept coefficient function $g_0(\cdot)$ are established when 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 $g_j(\cdot)$ ($j \geq 1$) of endogenous variables are given only for the case when instrumental variables are the nearly weak case ($\alpha < 1/2$). More importantly, we show 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, for the cases when instrumental variables are weak as the case considered by Staiger and Stock (1997) ($\alpha = 1/2$), and 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 theoretical findings are 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 The setup

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

$$y_i = g_0(\mathbf{z}_{i1}) + \mathbf{g}(\mathbf{z}_{i1})' \mathbf{x}_i + u_i, \quad \mathbf{x}_i = n^{-\alpha} \mathbf{C}' \mathbf{z}_i + \mathbf{v}_i, \quad 1 \leq i \leq n,$$

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

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 \mathbf{z}_i . It is easy to show that

$$(3) \quad E(y_i | \mathbf{z}_i) = g_0(\mathbf{z}_{i1}) + E[\mathbf{x}_i | \mathbf{z}_i]' \mathbf{g}(\mathbf{z}_{i1}) \\ = g_0(\mathbf{z}_{i1}) + n^{-\alpha} \mathbf{z}'_i \mathbf{C} \mathbf{g}(\mathbf{z}_{i1}) \equiv \boldsymbol{\pi}(\mathbf{z}_i)' \mathbf{g}^*(\mathbf{z}_{i1}),$$

where $\boldsymbol{\pi}(\mathbf{z}_i)' = (1, n^{-\alpha} \mathbf{z}'_i \mathbf{C})$ and $\mathbf{g}^*(\mathbf{z}_{i1})' = (g_0(\mathbf{z}_{i1}), \mathbf{g}(\mathbf{z}_{i1})')$, which implies that $\{g_j(\cdot)\}$ are functional coefficients of $\boldsymbol{\pi}(\mathbf{z}_i)$, and $\{g_j(\cdot)\}$ could be estimated by running a nonparametric regression of y_i versus $\boldsymbol{\pi}(\mathbf{z}_i)$ if $\boldsymbol{\pi}(\mathbf{z}_i)$ were known. However, $\boldsymbol{\pi}(\mathbf{z}_i)$ is unknown in practice. Therefore, estimating $\{g_j(\cdot)\}$ requires a two-stage method. A preliminary step is to estimate $\boldsymbol{\pi}(\mathbf{z}_i)$ by regressing \mathbf{x}_i on \mathbf{z}_i , while the next step is to estimate $\{g_j(\cdot)\}$ by a regression of y_i on \mathbf{z}_i and the estimated value $\hat{\boldsymbol{\pi}}(\mathbf{z}_i)$ (the estimator of $\boldsymbol{\pi}(\mathbf{z}_i)$) at the first step. 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. Finally, note that when $\alpha = 0$, model (1) becomes the functional-coefficient IV model studied by Cai et al. (2006) with nonparametric reduced form in (2) and if there is no endogeneity ($E(u_i|\mathbf{x}_i, \mathbf{z}_{i1}) = 0$), it reduces to the ordinary functional-coefficient regression model explored by many authors, including Hastie and Tibshirani (1993) for iid data, Cai, Fan and Yao (2000) for time series and Cai and Xu (2008) for quantile regression.

For simplicity of presentation, we provide some additional definitions and notations. If \mathbf{W} is a $p \times q$ matrix, $\text{Vec}(\mathbf{W})$ denotes the $pq \times 1$ vector formed by stacking the columns of \mathbf{W} under each; that is, if $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q)$, where \mathbf{W}_i is a $p \times 1$ vector for $i = 1, \dots, q$, then $\text{Vec}(\mathbf{W}) = (\mathbf{W}'_1, \mathbf{W}'_2, \dots, \mathbf{W}'_q)'$. 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 when $k = 1$ in (1). Extension to the case when $k > 1$ 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 \mathbf{z}_{i1} to $z_{i1} \in \mathfrak{R}$ throughout this paper.

2.2 A two-stage estimator

Given observed data $\{(y_i, \mathbf{x}_i, \mathbf{z}_i)\}_{i=1}^n$, our suggested estimation procedure is a two-stage approach, described as follows. The first stage involves the estimation of $\boldsymbol{\pi}(\mathbf{z}_i)$ by using least squares estimation to model (2) and the second stage is to use a local linear regression to model (3) by replacing $\boldsymbol{\pi}(\mathbf{z}_i)$ in (3) by the estimated value of $\boldsymbol{\pi}(\mathbf{z}_i)$, denoted by $\hat{\boldsymbol{\pi}}(\mathbf{z}_i)$.

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

$$(4) \quad \mathbf{x} = n^{-\alpha} \mathbf{z} \mathbf{C} + \mathbf{v},$$

where $\mathbf{x}' = (\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n)$, $\mathbf{z}' = (\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_n)$, and $\mathbf{v}' = (\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n)$. Then, using the least squares estimation to reduced form equation (4), we have

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

Now, we derive the local linear estimator of $\{g_j(\cdot)\}$. For this purpose, we assume throughout this paper that the functions $\{g_j(\cdot)\}$ have a continuous second partial derivative at any given point $z_1 \in \mathfrak{R}$. 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 $\theta_{2,j} = g_j^{(1)}(z_1) = \partial g_j(z_1)/\partial z_1$. Denote $\boldsymbol{\pi}_i = \boldsymbol{\pi}(\mathbf{z}_i)$, and $\hat{\boldsymbol{\pi}}_i = \hat{\boldsymbol{\pi}}(\mathbf{z}_i) = (1, n^{-\alpha} \mathbf{z}'_i \hat{\mathbf{C}})'$ as well as $\boldsymbol{\Pi}'_i = (\boldsymbol{\pi}'_i(z_{i1} - z_1) \boldsymbol{\pi}'_i)$. Then, the conditional mean in model (3) can be approximated by $E(y_i | \mathbf{z}_i) \approx \boldsymbol{\Pi}'_i \boldsymbol{\Theta}$, where $\boldsymbol{\Theta} = \boldsymbol{\Theta}(z_1) = (\boldsymbol{\theta}'_1 \ \boldsymbol{\theta}'_2)'$, $\boldsymbol{\theta}_1 = (\theta_{1,0} \ \dots \ \theta_{1,p})'$, and $\boldsymbol{\theta}_2 = (\theta_{2,0} \ \dots \ \theta_{2,p})'$. The local linear estimator $\hat{\boldsymbol{\Theta}}$ is defined as the minimizer of the following sum of locally weighted least squares

$$(6) \quad \sum_{i=1}^n \left[y_i - \sum_{j=0}^p \{\theta_{1,j} + (z_{i1} - z_1)\theta_{2,j}\} \hat{\pi}_{i,j} \right]^2 K_h(z_{i1} - z_1) \\ = \sum_{i=1}^n \left[y_i - \hat{\boldsymbol{\Pi}}_i' \boldsymbol{\Theta} \right]^2 K_h(z_{i1} - z_1),$$

where $\hat{\pi}_{i,j}$ denotes the j -th element of $\hat{\boldsymbol{\pi}}_i$, $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a kernel function on \mathfrak{R} , $h > 0$ is the bandwidth at the second step, $h \rightarrow 0$ and $nh \rightarrow \infty$. By minimizing (6) with respect to $\boldsymbol{\Theta}$, 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 $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\Pi}} \mathbf{W} \hat{\boldsymbol{\Pi}}')^{-1} \hat{\boldsymbol{\Pi}} \mathbf{W} \mathbf{Y}$, where $\mathbf{Y} = (y_1, y_2, \dots, y_n)'$, $\mathbf{W} = \text{diag}\{K_h(z_{11} - z_1), \dots, K_h(z_{n1} - z_1)\}$, and $\hat{\boldsymbol{\Pi}} = (\hat{\boldsymbol{\Pi}}_1 \ \dots \ \hat{\boldsymbol{\Pi}}_n)$ is the estimator of $\boldsymbol{\Pi} = (\boldsymbol{\Pi}_1 \ \dots \ \boldsymbol{\Pi}_n)$. It is easily verified that $\hat{\boldsymbol{\Theta}}$ can be re-written as $\hat{\boldsymbol{\Theta}} = \mathbf{H}^{-1} \hat{\mathbf{S}}_n^{-1} \hat{\mathbf{T}}_n$, where $\mathbf{H} = \mathbf{H}(h) = \text{diag}\{\mathbf{I}_{p+1}, h \mathbf{I}_{p+1}\}$, \mathbf{I}_{p+1} is the $(p+1) \times (p+1)$ identity matrix,

$$(7) \quad \hat{\mathbf{S}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_{i1}^{*\otimes 2} \otimes \hat{\boldsymbol{\pi}}_i^{*\otimes 2} K_h(z_{i1} - z_1), \quad \text{and} \\ \hat{\mathbf{T}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_{i1}^* \otimes \hat{\boldsymbol{\pi}}_i K_h(z_{i1} - z_1) y_i$$

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

2.3 Distribution theory

2.3.1. Assumptions and notations

Set $\mathbf{H}_1 = \mathbf{H}_1(n) = \text{diag}\{1, n^\alpha \mathbf{I}_{p+1}\}$, $\mathbf{H}_2 = \mathbf{H}_2(n) = \text{diag}\{1, n^{1/2} \mathbf{I}_{p+1}\}$, and $\mathbf{H}_3 = \mathbf{H}_3(n) = \mathbf{H}_2$. Let $f_1(\cdot)$ be the probability density function of z_{i1} . Define $\mu_2(K) = \int u^2 K(u) du$ and $\nu_j(K) = \int u^j K^2(u) du$. The following conditions are listed for the asymptotic theory.

Assumptions:

- A1. The kernel $K(\cdot)$ is symmetric and bounded second order kernel function.
- A2. $\{\mathbf{z}_i\}$ are independent and identically distributed. $\boldsymbol{\Sigma}_{z z'} = E(\mathbf{z}_i \mathbf{z}'_i)$ exists and is positive definite. Also, the conditional covariance matrix of \mathbf{z}_i given $z_{i1} = z_1$, $\mathbf{M}_2(z_1) = E[\mathbf{z}_i \mathbf{z}'_i | z_{i1} = z_1]$ is positive definite for a given grid point z_1 .

- A3. The second order derivative functions $\{g_j^{(2)}(z_1)\}$ are continuous at a given grid point z_1 .
- A4. $\{(u_i, \mathbf{v}_i)'\}$ are independent and identically distributed with the mean zero and conditional covariance matrix of $(u_i, \mathbf{v}_i)'$ given \mathbf{z}_i is $\Sigma = \Sigma(\mathbf{z}_i) = \begin{pmatrix} \sigma_{uu}(\mathbf{z}_i) & \Sigma_{uv}(\mathbf{z}_i) \\ \Sigma_{vu}(\mathbf{z}_i) & \Sigma_{vv}(\mathbf{z}_i) \end{pmatrix}$, positive definite for all \mathbf{z}_i .
- A5. $h \rightarrow 0$ and $nh \rightarrow \infty$.
- A6. The density function $f_1(\cdot)$ is continuous and $f_1(z_1) > 0$ at a given grid point z_1 .
- A7. $E((z_{ij_1}, z_{ij_2})^2) < \infty$ for all $1 \leq j_1, j_2 \leq (1+q)$.

Assumptions A1–A3 and A5–A6 are commonly imposed in local polynomial smoothing methods; see Fan and Gijbels (1996). The asymptotic sampling theory for the resulting two-stage estimator 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 $\mathbf{M}_1(z_1) = E(\mathbf{z}_i | z_{i1} = z_1)$ and $\mathbf{M}(z_1) = E[\mathbf{z}_i^* \otimes 2 | z_{i1} = z_1]$, where $\mathbf{z}_i^* = (1 \quad \mathbf{z}_i')$. Set $\Delta_{uu} = E[\sigma_{uu}(\mathbf{z}_i) \mathbf{z}_i^* \otimes 2 | z_{i1} = z_1]$, $\Delta_{vu} = \mathbf{g}(z_1)' E \times [\Sigma_{vu}(\mathbf{z}_i) \mathbf{z}_i^* \otimes 2 | z_{i1} = z_1]$, $\Delta_{uv} = \Delta_{vu}' = \Delta_{vu}$, $\Delta_{vv} = E[\mathbf{g}(z_1)' \Sigma_{vv}(\mathbf{z}_i) \mathbf{g}(z_1) \mathbf{z}_i^* \otimes 2 | z_{i1} = z_1]$, and $\Lambda = \text{diag}\{\Lambda_1, \Lambda_3\}$, where $\Lambda_3 = E[\Sigma_{vv}(\mathbf{z}_i) \otimes \mathbf{z}_i^* \otimes 2]$ and

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

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 $1 \leq j \leq 3$, $\mathbf{S}^{(j)}(z_1) = \text{diag}\{1, \mu_2(K)\} \otimes \Omega^{(j)}(z_1)$, where $\Omega^{(j)}(z_1) = \mathbf{D}_j^* \mathbf{M}(z_1) \mathbf{D}_j^*$, $\mathbf{D}_j^* = \text{diag}\{1, \mathbf{D}_j\}$, $\mathbf{D}_1 = \mathbf{C}$, $\mathbf{D}_2 = \mathbf{C} + \Sigma_{zz}^{-1} \mathbf{Z}_v$, and $\mathbf{D}_3 = \Sigma_{zz}^{-1} \mathbf{Z}_v$ with \mathbf{Z}_v being a $(q+1) \times p$ matrix of random variables and $\text{Vec}(\mathbf{Z}_v) \sim N(\mathbf{0}, \Lambda_3)$. Finally, define $\varepsilon_i = y_i - E(y_i | \mathbf{z}_i)$. Then, by (3), $\varepsilon_i = u_i + \mathbf{v}_i' \mathbf{g}(z_{i1})$ and $\Delta_\varepsilon = \text{Var}(\varepsilon_i \mathbf{z}_i^* | z_{i1} = z_1) = E[\sigma_\varepsilon^2(\mathbf{z}_i) \mathbf{z}_i^* \otimes 2 | z_{i1} = z_1] = \Delta_{uu} + 2\Delta_{uv} + \Delta_{vv}$, where $\sigma_\varepsilon^2(\mathbf{z}_i)$ is the conditional variance of ε_i given \mathbf{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 A1–A7, we have (i) for $c(\alpha) = 1$,*

$$\begin{aligned} & \sqrt{nh} (\mathbf{I}_2 \otimes \mathbf{H}_1^{-1}) \left[\mathbf{H} \left\{ \hat{\Theta} - \Theta \right\} - \frac{h^2}{2} \begin{pmatrix} \mu_2(K) \mathbf{g}^{*(2)}(z_1) \\ \mathbf{0} \end{pmatrix} \right] \\ & \Rightarrow \left[f_1(z_1) \mathbf{S}^{(1)}(z_1) \right]^{-1} (\mathbf{I}_2 \otimes \mathbf{D}_1^*(z_1)') (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

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

$$\begin{aligned} & \sqrt{nh} (\mathbf{I}_2 \otimes \mathbf{H}_2^{-1}) \left[\mathbf{H} \left\{ \hat{\Theta} - \Theta \right\} - \frac{h^2}{2} \begin{pmatrix} \mu_2(K) g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} \right] \\ & \Rightarrow \left[f_1(z_1) \mathbf{S}^{(c(\alpha))}(z_1) \right]^{-1} (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^*(z_1)') (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

where \mathbf{Z}_{ku} and \mathbf{Z}_{kvg} are $2(q+1) \times 1$ normal random vectors and the joint distribution of \mathbf{Z}_{ku} , \mathbf{Z}_{kvg} , and $\text{Vec}(\mathbf{Z}_v)$ is $N(\mathbf{0}, \Lambda)$.

Note that \mathbf{Z}_{ku} and \mathbf{Z}_{kvg} are independent of \mathbf{Z}_v . By Theorem 1, we are ready to have the asymptotic distributions of the estimators $\hat{g}_0(\cdot)$ and $\hat{\mathbf{g}}(\cdot)$, which are provided in Corollary 1.

Corollary 1. *Under Assumptions A1–A7, then, (i) for $c(\alpha) = 1$,*

$$\begin{aligned} & \sqrt{nh} \mathbf{H}_1^{-1} \begin{pmatrix} \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g_0^{(2)}(z_1) \\ \hat{\mathbf{g}}(z_1) - \mathbf{g}(z_1) - \frac{h^2}{2} \mu_2(K) \mathbf{g}^{(2)}(z_1) \end{pmatrix} \\ & \Rightarrow f_1^{-1}(z_1) \Omega^{(1)}(z_1)^{-1} \mathbf{D}_1^* (\mathbf{I}_{q+2}, \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

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

$$\begin{aligned} & \sqrt{nh} \mathbf{H}_2^{-1} \begin{pmatrix} \hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g_0^{(2)}(z_1) \\ \hat{\mathbf{g}}(z_1) - \mathbf{g}(z_1) \end{pmatrix} \\ & \Rightarrow f_1^{-1}(z_1) \Omega^{(c(\alpha))}(z_1)^{-1} \mathbf{D}_{c(\alpha)}^* (\mathbf{I}_{q+2}, \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

where \mathbf{Z}_{ku} and \mathbf{Z}_{kvg} are given in Theorem 1.

Corollary 2. *Under Assumptions A1–A7, (i) for $c(\alpha) = 1$,*

$$\begin{aligned} & \sqrt{nh} \left[\hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g_0^{(2)}(z_1) \right] \\ & \Rightarrow f_1^{-1}(z_1) (\Omega_1, -\mathbf{M}_1(z_1)' \mathbf{D}_1 \mathbf{J}_1 \mathbf{D}_1', \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

and

$$\begin{aligned} & n^{1/2-\alpha} h^{1/2} \left[\hat{\mathbf{g}}(z_1) - \mathbf{g}(z_1) - \frac{h^2}{2} \mu_2(K) \mathbf{g}^{(2)}(z_1) \right] \\ & \Rightarrow f_1^{-1}(z_1) (-\mathbf{J}_1 \mathbf{D}_1' \mathbf{M}_1(z_1), \mathbf{J}_1 \mathbf{D}_1', \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

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

$$\begin{aligned} & \sqrt{nh} \left[\hat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g_0^{(2)}(z_1) \right] \\ & \Rightarrow f_1^{-1}(z_1) (\Omega_{c(\alpha)}, -\mathbf{M}_1(z_1)' \mathbf{D}_{c(\alpha)} \mathbf{J}_{c(\alpha)} \mathbf{D}_{c(\alpha)}', \mathbf{0}) \\ & \quad \times (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

and

$$\begin{aligned} & h^{1/2} [\hat{\mathbf{g}}(z_1) - \mathbf{g}(z_1)] \\ & \Rightarrow f_1^{-1}(z_1) (-\mathbf{J}_{c(\alpha)} \mathbf{D}_{c(\alpha)}' \mathbf{M}_1(z_1), \mathbf{J}_{c(\alpha)} \mathbf{D}_{c(\alpha)}', \mathbf{0}) \\ & \quad \times (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

where for $1 \leq j \leq 3$, $\mathbf{J}_j = [\mathbf{D}'_j(\mathbf{M}_2(z_1) - \mathbf{M}_1(z_1) \times \mathbf{M}_1(z_1)') \mathbf{D}_j]^{-1}$, and $\Omega_j = 1 + \mathbf{M}_1(z_1)' \mathbf{D}_j \mathbf{J}_j \mathbf{D}'_j \mathbf{M}_1(z_1)$.

Remark 1. It follows clearly from Corollary 2 that the two-stage estimator for $g_0(\cdot)$ is always consistent with the same convergence rate at \sqrt{nh} for any α although the magnitudes might be different for different values of α . However, the two-stage estimator for $\mathbf{g}(\cdot)$ is consistent only for the nearly weak case $0 < \alpha < 1/2$ and the rate of convergence is lower than that for $\widehat{g}_0(\cdot)$. 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 $\widehat{g}_0(z_1)$ and $\widehat{\mathbf{g}}(\cdot)$ for the case $0 < \alpha < 1/2$, stated in Corollary 3. These findings are novel in the literature.

Corollary 3. Under Assumptions A1–A7, for $0 < \alpha < 1/2$, if $nh^5 = O(1)$, then

$$\sqrt{nh} \left[\widehat{g}_0(z_1) - g_0(z_1) - \frac{h^2}{2} \mu_2(K) g_0^{(2)}(z_1) \right] \Rightarrow N(0, \sigma_{g_0}^2(z_1)),$$

and if $n^{1-2\alpha}h \rightarrow \infty$ and $n^{1-2\alpha}h^5 = O(1)$, then

$$n^{1/2-\alpha} h^{1/2} \left[\widehat{\mathbf{g}}(z_1) - \mathbf{g}(z_1) - \frac{h^2}{2} \mu_2(K) \mathbf{g}^{(2)}(z_1) \right] \Rightarrow N(\mathbf{0}, \boldsymbol{\Sigma}_g(z_1)),$$

where $\sigma_{g_0}^2(z_1) = f_1^{-1}(z_1) \nu_0(K) \mathbf{e}'_1 [\boldsymbol{\Omega}^{(1)}(z_1)]^{-1} \mathbf{D}_1^* \boldsymbol{\Delta}_\varepsilon \mathbf{D}_1^* \times [\boldsymbol{\Omega}^{(1)}(z_1)]^{-1} \mathbf{e}_1$, $\mathbf{e}_1 = (1, 0, \dots, 0)'$, $\boldsymbol{\Sigma}_g(z_1) = f_1^{-1}(z_1) \nu_0(K) \times \mathbf{e}'_2 [\boldsymbol{\Omega}^{(1)}(z_1)]^{-1} \mathbf{D}_1^* \boldsymbol{\Delta}_\varepsilon \mathbf{D}_1^* [\boldsymbol{\Omega}^{(1)}(z_1)]^{-1} \mathbf{e}_2$, and $\mathbf{e}'_2 = (0, \mathbf{I}_p)$.

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 $\widehat{g}_0(z_1)$, which is

$$\begin{aligned} \text{AMSE}_0 &= \text{AMSE}(\widehat{g}_0(z_1)) \\ &= \frac{1}{4} h^4 \mu_2^2(K) \left[g_0^{(2)}(z_1) \right]^2 + \sigma_{g_0}^2(z_1)/(nh). \end{aligned}$$

By minimizing AMSE_0 with respect to h , we obtain the optimal bandwidth for $\widehat{g}_0(z_1)$, which is $h_{0,opt} = O(n^{-1/5})$, so that the optimal AMSE_0 has an order $O(n^{-4/5})$. Similarly, for $j \geq 1$, the optimal bandwidth for estimating $g_j(z_1)$ is $h_{j,opt} = O(n^{-(1-2\alpha)/5})$ and the optimal AMSE for $\widehat{g}_j(z_1)$ is $O(n^{-4/5+8\alpha/5})$, 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 $\mathbf{g}(\cdot)$ optimally in the sense of minimizing the asymptotic mean squared error. Therefore, to estimate both $g_0(\cdot)$ and $\mathbf{g}(\cdot)$ 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 future studies.

Remark 3. There is an important practical aspect to be noted. The question is now how to adaptively select the bandwidth in practice. Here, we suggest rule-of-thumb bandwidths. Based on the aforementioned discussions and results, to estimate $g_0(\cdot)$ for all three cases, the choice of bandwidth can be carried out as in standard nonparametric regression. In that case, a number of methods could be used to select the bandwidth such as the plug-in approach in Ruppert, Sheather and Wand (1995); see Cai et al. (2006) for further discussions. To estimate $\mathbf{g}(\cdot)$ for the case when $c(\alpha) = 1$, first, use the bandwidth selected earlier to estimate all coefficient functions and then obtain the partial residual $y_i^* = y_i - \widehat{g}_0(\mathbf{z}_{i1})$. Finally, estimate $\mathbf{g}(\cdot)$ through the regression model $y_i^* = \mathbf{g}(\mathbf{z}_{i1})' \mathbf{x}_i + u_i$ by using a bandwidth selected based on the aforementioned methods. However, there appears to be no results available in the literature for a data-driven bandwidth selection with optimal properties for a two-stage nonparametric estimation procedure; see Newey, Powell and Vella (1999) for the related discussion. It is 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 simulation studies (see Section 3).

Finally, we remark from Corollary 3 that the asymptotic variance (sandwich form) consists of three terms: the first term $\boldsymbol{\Delta}_{uu}$ in the meat part $\boldsymbol{\Delta}_\varepsilon$ addresses the variation of measurement error at the second step, the second term $\boldsymbol{\Delta}_{uv}$ accounts correctly for the asymptotic covariance between the first and second steps, and the third term $\boldsymbol{\Delta}_{vv}$ 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, Fang 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 and the ad hoc bandwidth selection approach mentioned in Remark 3 is used in our simulations. We evaluate the finite sample performances of our estimator in terms of the mean absolute deviation error (MADE)

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

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

We consider the following data generating model,

$$y_i = g_0(z_{i1}) + g_1(z_{i2}) x_i + u_i, \quad x_i = 2n^{-\alpha} z_{i1} - 3n^{-\alpha} z_{i2} + v_i,$$

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

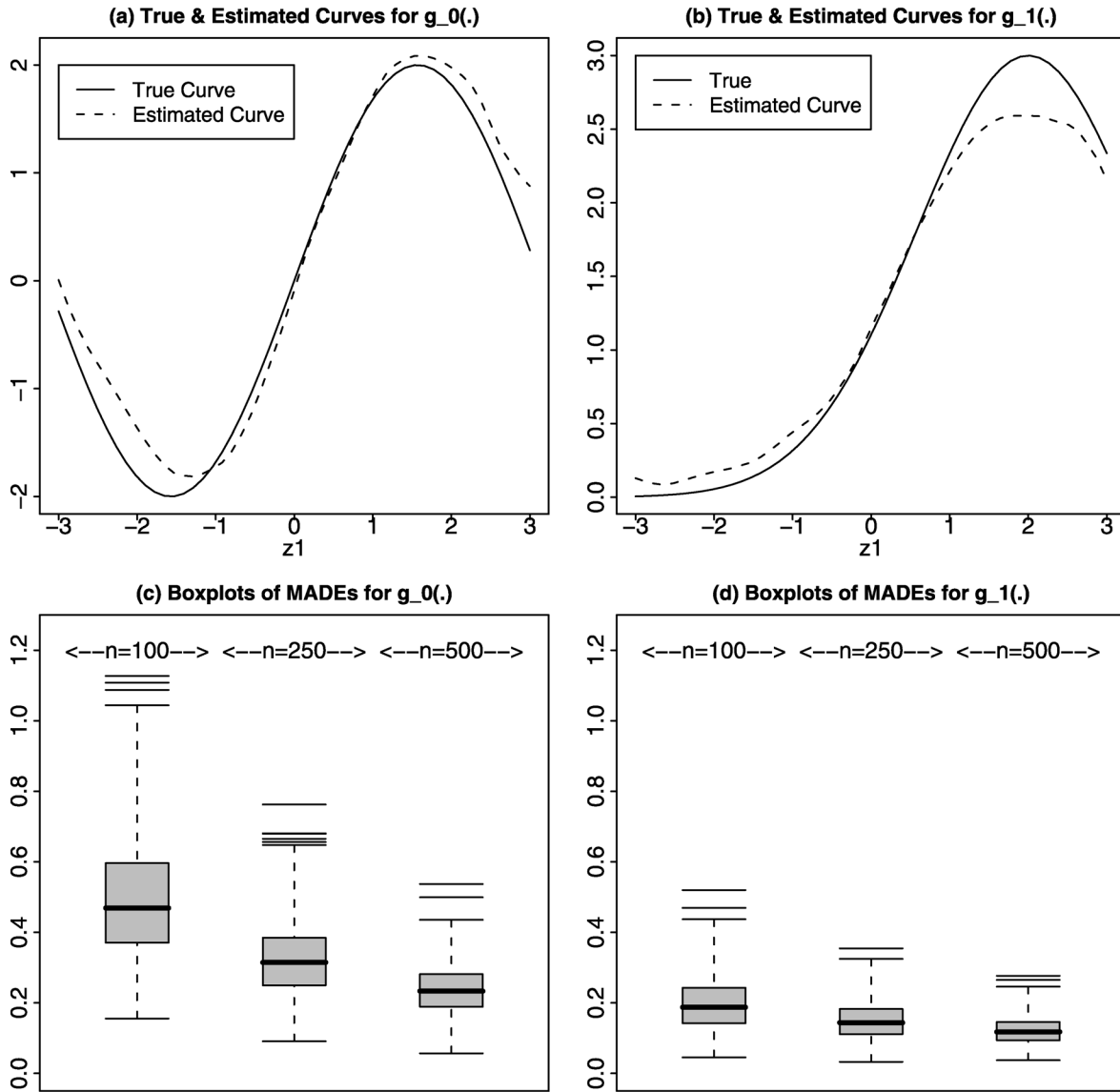


Figure 1. Simulation results for Case I ($\alpha = 0.2$). (a) The true coefficient function $g_0(\cdot)$ (solid line) with the two-stage local linear estimator (dashed line). (b) The true coefficient function $g_1(\cdot)$ (solid line) with the two-stage local linear estimator (dashed line). (c) Boxplots of the 500 MADE values of $\hat{g}_0(\cdot)$. (d) Boxplots of the 500 MADE values of $\hat{g}_1(\cdot)$.

$(-3, 3)$, 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: $n = 100, 250$, and 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, the MADE values decrease 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 ($= \mathcal{E}_0 + \mathcal{E}_1$)

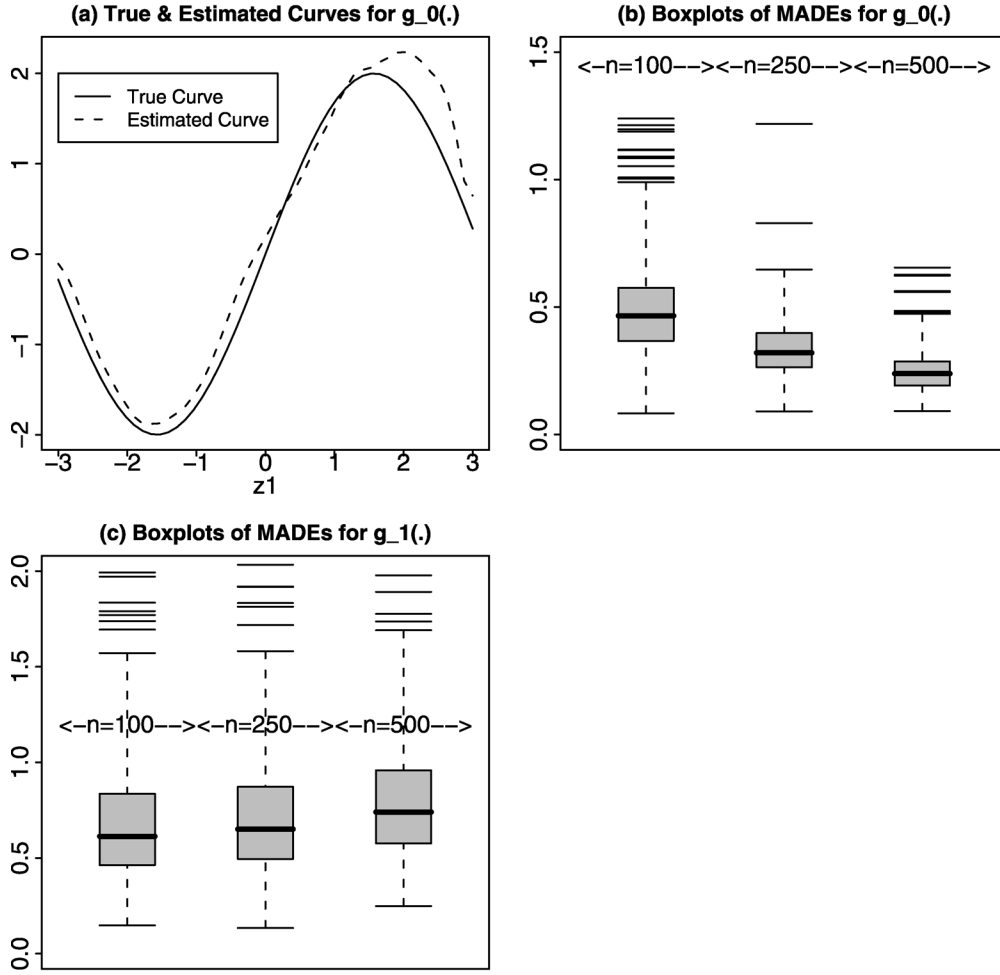


Figure 2. Simulation results for Case II ($\alpha = 0.5$). (a) The true coefficient function $g_0(\cdot)$ (solid line) and its two-stage local linear estimator (dashed line). (b) Boxplots of the 500 MADE values of $\hat{g}_0(\cdot)$. (c) Boxplots of the 500 MADE values of $\hat{g}_1(\cdot)$.

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 $\hat{g}_0(\cdot)$ is consistent. But the MADE for $\hat{g}_1(\cdot)$ 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 ($= \mathcal{E}_0$) 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. MATHEMATICAL 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 $\hat{\mathbf{S}}_n$ 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 A1–A7, then

$$(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}) \hat{\mathbf{S}}_n (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}) \Rightarrow f_1(z_1) \mathbf{S}^{(c(\alpha))}(z_1),$$

where $\mathbf{S}^{(j)}(z_1)$ is given in Section 2.

Before we embrace on establishing the asymptotic properties of the resulting estimator, first, we decompose $\mathbf{H}[\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}]$

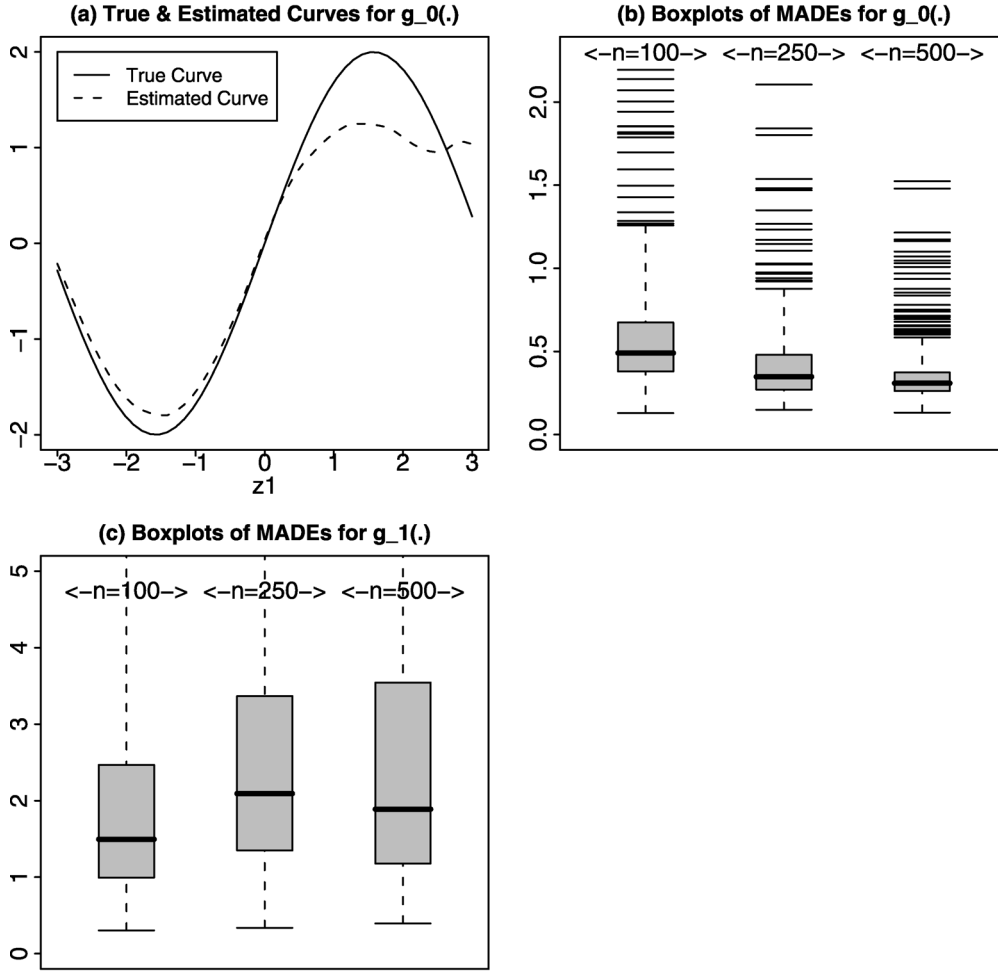


Figure 3. Simulation results of Case III ($\alpha = 0.7$). Caption is the same as that for Case II.

into three terms as

$$(8) \quad \mathbf{H} [\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}] \equiv \hat{\mathbf{S}}_n^{-1} [\mathbf{P}_n + \mathbf{Q}_n + \mathbf{R}_n],$$

where with $\mathbf{G} = (\boldsymbol{\pi}'_1 \mathbf{g}^*(z_{11}), \dots, \boldsymbol{\pi}'_n \mathbf{g}^*(z_{n1}))'$, $\mathbf{P}_n = n^{-1} \mathbf{H}^{-1} \hat{\boldsymbol{\Pi}} \mathbf{W} (\mathbf{Y} - \mathbf{G})$, $\mathbf{Q}_n = n^{-1} \mathbf{H}^{-1} \hat{\boldsymbol{\Pi}} \mathbf{W} (\mathbf{G} - \boldsymbol{\Pi}' \boldsymbol{\Theta})$, and $\mathbf{R}_n = n^{-1} \mathbf{H}^{-1} \hat{\boldsymbol{\Pi}} \mathbf{W} (\boldsymbol{\Pi}' \boldsymbol{\Theta} - \hat{\boldsymbol{\Pi}}' \boldsymbol{\Theta})$. The reason of doing the above decomposition is to show that \mathbf{P}_n contributes to only the asymptotic variance, \mathbf{Q}_n is the resource of the bias, and \mathbf{R}_n is negligible comparing with \mathbf{P}_n , which are presented in the following lemmas.

Lemma 2. Under Assumptions A1–A6, we have

$$\sqrt{nh} (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}) \mathbf{P}_n \Rightarrow (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^*) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}),$$

where the distributions of \mathbf{Z}_{ku} and \mathbf{Z}_{kvg} are given in Theorem 1.

Lemma 3. Under Assumptions A1–A7, we have

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

$$(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}) \mathbf{Q}_n = f_1(z_1) \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \otimes (\mathbf{D}_{c(\alpha)}^* \mathbf{M}(z_1) \mathbf{D}_1^*) \\ \times \mathbf{H}_{c(\alpha)}^{-1} \mathbf{g}^{*(2)}(z_1) \left[\frac{h^2}{2} + o_p(h^2) \right],$$

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

$$(\mathbf{I}_2 \otimes \mathbf{H}_3) \mathbf{Q}_n = f_1(z_1) \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \otimes (\mathbf{D}_3^* \mathbf{M}(z_1) \mathbf{D}_1^*) \\ \times \mathbf{H}_1^{-1} \mathbf{g}^{*(2)}(z_1) \left[\frac{h^2}{2} + o_p(h^2) \right].$$

Next, we consider the term \mathbf{R}_n , which is decomposed into two terms as $\mathbf{R}_n \equiv \mathbf{R}_{n,1} + \mathbf{R}_{n,2}$, where $\mathbf{R}_{n,1} = n^{-\alpha-1} \sum_{i=1}^n \mathbf{z}_{i1}^* \otimes \hat{\boldsymbol{\pi}}_i K_h(z_{i1} - z_1) \mathbf{z}'_i (\mathbf{C} - \hat{\mathbf{C}}) \mathbf{g}(z_1)$, and $\mathbf{R}_{n,2} = n^{-\alpha-1} \sum_{i=1}^n \mathbf{z}_{i1}^* \otimes \hat{\boldsymbol{\pi}}_i K_h(z_{i1} - z_1) \mathbf{z}'_i (z_{i1} - z_1) (\mathbf{C} - \hat{\mathbf{C}}) \mathbf{g}'(z_1)$. Then, we have the following results for both $\mathbf{R}_{n,1}$ and $\mathbf{R}_{n,2}$, respectively.

Lemma 4. Under Assumptions A1–A7, we have

$$\begin{aligned} n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{R}_{n,1} &= O_p(1), \quad \text{and} \\ h^{-1}n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{R}_{n,2} &= O_p(1). \end{aligned}$$

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

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

$$(9) \quad \mathbf{H} \left[\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right] - \widehat{\mathbf{S}}_n^{-1} \mathbf{Q}_n - \widehat{\mathbf{S}}_n^{-1} \mathbf{R}_n = \widehat{\mathbf{S}}_n^{-1} \mathbf{P}_n.$$

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

$$\begin{aligned} \sqrt{n}h(\mathbf{I}_2 \otimes \mathbf{H}_1^{-1})\widehat{\mathbf{S}}_n^{-1} \mathbf{R}_n &= h^{1/2} \left[(\mathbf{I}_2 \otimes \mathbf{H}_1)\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_1) \right]^{-1} \\ &\quad \times n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_1(n))\mathbf{R}_n \rightarrow^p \mathbf{0}, \end{aligned}$$

and by Lemmas 1 and 2,

$$\begin{aligned} \sqrt{n}h(\mathbf{I}_2 \otimes \mathbf{H}_1^{-1})\widehat{\mathbf{S}}_n^{-1} \mathbf{P}_n &= \left[(\mathbf{I}_2 \otimes \mathbf{H}_1)\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_1) \right]^{-1} \\ &\quad \times n^{1/2}h^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_1)\mathbf{P}_n \\ &\Rightarrow \left[f_1(z_1)\mathbf{S}^{(1)}(z_1) \right]^{-1} \left(\mathbf{I}_2 \otimes \mathbf{D}_1^{*'} \right) \\ &\quad \times (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}). \end{aligned}$$

By Lemmas 1 and 3 and the fact that $\mathbf{H}_1^{-1}(n)\mathbf{g}^{*(2)}(z_1) = \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o(1)$,

$$\begin{aligned} &(\mathbf{I}_2 \otimes \mathbf{H}_1^{-1}(n))\widehat{\mathbf{S}}_n^{-1} \mathbf{Q}_n \\ &= \left[(\mathbf{I}_2 \otimes \mathbf{H}_1)\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_1) \right]^{-1} (\mathbf{I}_2 \otimes \mathbf{H}_1(n))\mathbf{Q}_n \\ &= \frac{h^2}{2} \left[f_1(z_1)\mathbf{S}^{(1)}(z_1) \right]^{-1} f_1(z_1) \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \\ &\quad \otimes \boldsymbol{\Omega}^{(1)}(z_1)\mathbf{H}_1^{-1}\mathbf{g}^{*(2)}(z_1) + o_p(h^2) \\ &= \frac{h^2}{2} \begin{pmatrix} \mu_2(K)\mathbf{H}_1^{-1}(n)\mathbf{g}^{*(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o_p(h^2) \\ &= \frac{h^2}{2} \begin{pmatrix} \mu_2(K)g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o_p(h^2). \end{aligned}$$

Thus,

$$\begin{aligned} &\sqrt{n}h(\mathbf{I}_2 \otimes \mathbf{H}_1^{-1}) \left[\mathbf{H} \left\{ \widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right\} - \frac{h^2}{2} \begin{pmatrix} \mu_2(K)\mathbf{g}^{*(2)}(z_1) \\ \mathbf{0} \end{pmatrix} \right] \\ &\Rightarrow \left[f_1(z_1)\mathbf{S}^{(1)}(z_1) \right]^{-1} \left(\mathbf{I}_2 \otimes \mathbf{D}_1^{*'} \right) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}). \end{aligned}$$

Second, we consider the weak case ($\alpha = 1/2$). Similar to the above arguments, we have

$$\sqrt{n}h(\mathbf{I}_2 \otimes \mathbf{H}_2^{-1})\widehat{\mathbf{S}}_n^{-1} \mathbf{R}_n \rightarrow^p \mathbf{0},$$

$$\begin{aligned} &\sqrt{n}h(\mathbf{I}_2 \otimes \mathbf{H}_2^{-1})\widehat{\mathbf{S}}_n^{-1} \mathbf{P}_n \\ &\Rightarrow \left[f_1(z_1)\mathbf{S}^{(2)}(z_1) \right]^{-1} \left(\mathbf{I}_2 \otimes \mathbf{D}_2^{*'} \right) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}), \end{aligned}$$

and

$$\begin{aligned} (\mathbf{I}_2 \otimes \mathbf{H}_2^{-1})\widehat{\mathbf{S}}_n^{-1} \mathbf{Q}_n &= \frac{h^2}{2} \left[\mathbf{S}^{(2)}(z_1) \right]^{-1} \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \\ &\quad \otimes (\mathbf{D}_2^{*'} \mathbf{M}(z_1) \mathbf{D}_1^* \mathbf{H}_2^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2)) \\ &\equiv \mathbf{K}_n + o_p(h^2), \end{aligned}$$

where

$$\begin{aligned} \mathbf{K}_n &= \frac{h^2}{2} \left[\mathbf{S}^{(2)}(z_1) \right]^{-1} \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \\ &\quad \otimes \left[\mathbf{D}_2^{*'} \mathbf{M}(z_1) \mathbf{D}_1^* \mathbf{H}_2^{-1} \mathbf{g}^{*(2)}(z_1) \right]. \end{aligned}$$

Then, by (9),

$$\begin{aligned} &\sqrt{n}h \left[(\mathbf{I}_2 \otimes \mathbf{H}_2^{-1})\mathbf{H} \left\{ \widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right\} - \mathbf{K}_n + o_p(h^2) \right] \\ &\Rightarrow \left[f_1(z_1)\mathbf{S}^{(2)}(z_1) \right]^{-1} \left(\mathbf{I}_2 \otimes \mathbf{D}_2^{*'} \right) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}). \end{aligned}$$

Now, we evaluate the term \mathbf{K}_n . By definitions of \mathbf{D}_1^* and \mathbf{D}_2^* , it is clear that

$$\begin{aligned} &\mathbf{D}_2^{*'} \mathbf{M}(z_1) \mathbf{D}_1^* \mathbf{H}_2^{-1} \mathbf{g}^{*(2)}(z_1) \\ &= \mathbf{D}_2^{*'} \mathbf{M}(z_1) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 \end{pmatrix} \begin{pmatrix} g_0^{(2)}(z_1) \\ n^{-1/2} \mathbf{g}^{(2)}(z_1) \end{pmatrix} \\ &= \mathbf{D}_2^{*'} \mathbf{M}(z_1) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 \end{pmatrix} \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o(1) \\ &= \mathbf{D}_2^{*'} \mathbf{M}(z_1) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o(1) \\ &= \mathbf{D}_2^{*'} \mathbf{M}(z_1) \mathbf{D}_2^* \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o(1) \\ &= \boldsymbol{\Omega}^{(2)}(z_1) \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o(1). \end{aligned}$$

Therefore,

$$\begin{aligned} &\begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \otimes \left[\mathbf{D}_2^{*'} \mathbf{M}(z_1) \mathbf{D}_1^* \mathbf{H}_2^{-1} \mathbf{g}^{*(2)}(z_1) \right] \\ &= \begin{pmatrix} \mu_2(K)\boldsymbol{\Omega}^{(2)}(z_1) \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} + o(1). \end{aligned}$$

Hence, by the definition of $\mathbf{S}^{(2)}(z_1)$,

$$\begin{aligned} \mathbf{K}_n &= \frac{h^2}{2} \left[\mathbf{S}^{(2)}(z_1) \right]^{-1} \begin{pmatrix} \mu_2(K)\boldsymbol{\Omega}^{(2)}(z_1) \begin{pmatrix} g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} \\ \mathbf{0} \end{pmatrix} + o(h^2) \\ &= \frac{h^2}{2} \left[\begin{pmatrix} \mu_2(K)g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o_p(1) \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} & \sqrt{nh}(\mathbf{I}_2 \otimes \mathbf{H}_2^{-1}) \\ & \times \left[\mathbf{H} \left\{ \hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta} \right\} - \frac{h^2}{2} \begin{pmatrix} \mu_2(K)g_0^{(2)}(z_1) \\ \mathbf{0} \end{pmatrix} + o_p(h^2) \right] \\ & \Rightarrow \left[f_1(z_1)\mathbf{S}^{(2)}(z_1) \right]^{-1} \left(\mathbf{I}_2 \otimes \mathbf{D}_2^{*'} \right) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}). \end{aligned}$$

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

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

$$\begin{aligned} & \left[f_1(z_1)\mathbf{S}^{(c(\alpha))}(z_1) \right]^{-1} (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^{*'}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ & = \left[f_1^{-1}(z_1) \begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix}^{-1} \otimes \boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} \right] \\ & \quad \times (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^{*'}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ & = f_1^{-1}(z_1) \begin{pmatrix} \boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} & \mathbf{0} \\ \mathbf{0} & \mu_2^{-1}(K)\boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} \end{pmatrix} \\ & \quad \times \begin{pmatrix} \mathbf{D}_{c(\alpha)}^{*'} (\mathbf{I}_{q+2} & \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ \mathbf{D}_{c(\alpha)}^{*'} (\mathbf{0} & \mathbf{I}_{q+2}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \end{pmatrix} \\ & = f_1^{-1}(z_1) \\ & \quad \times \begin{pmatrix} \boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} \mathbf{D}_{c(\alpha)}^{*'} (\mathbf{I}_{q+2} & \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ \mu_2^{-1}(K)\boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} \mathbf{D}_{c(\alpha)}^{*'} (\mathbf{0} & \mathbf{I}_{q+2}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \end{pmatrix}. \end{aligned}$$

Therefore, the corollary holds. \square

Proof of Corollary 2. By the inverse of a partitioned matrix,

$$\begin{aligned} & \boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1} \\ & = \begin{pmatrix} \boldsymbol{\Omega}_{c(\alpha)} & -\mathbf{M}_1(z_1)'\mathbf{D}_{c(\alpha)}\mathbf{J}_{c(\alpha)} \\ -\mathbf{J}_{c(\alpha)}\mathbf{D}'_{c(\alpha)}\mathbf{M}_1(z_1) & \mathbf{J}_{c(\alpha)} \end{pmatrix}, \end{aligned}$$

which implies that

$$\begin{aligned} & f_1^{-1}(z_1)\boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1}\mathbf{D}_{c(\alpha)}^{*'} (\mathbf{I}_{q+2} \quad \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ & = f_1^{-1}(z_1)\boldsymbol{\Omega}^{(c(\alpha))}(z_1)^{-1}\mathbf{D}_{c(\alpha)}^{*'} \begin{pmatrix} (1 \quad \mathbf{0} \quad \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ (0 \quad \mathbf{I}_{q+1} \quad \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \end{pmatrix} \\ & = \begin{pmatrix} f_1^{-1}(z_1) (\boldsymbol{\Omega}_{c(\alpha)} & -\mathbf{M}_1(z_1)'\mathbf{D}_{c(\alpha)}\mathbf{J}_{c(\alpha)}\mathbf{D}'_{c(\alpha)} & \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \\ f_1^{-1}(z_1) (-\mathbf{J}_{c(\alpha)}\mathbf{D}'_{c(\alpha)}\mathbf{M}_1(z_1) & \mathbf{J}_{c(\alpha)}\mathbf{D}'_{c(\alpha)} & \mathbf{0}) (\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}) \end{pmatrix}. \end{aligned}$$

By Corollary 1, we have proved Corollary 2. \square

Proof of Corollary 3. It is easy to see from Lemma A.1 in Appendix that $\begin{pmatrix} \mathbf{Z}_{ku} \\ \mathbf{Z}_{kvg} \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Lambda}_1)$, and $(\mathbf{I}_{q+2} \quad \mathbf{0}) (\mathbf{Z}_{ku} +$

$\mathbf{Z}_{kvg}) \sim N(\mathbf{0}, \boldsymbol{\Lambda}_4)$, where $\boldsymbol{\Lambda}_4 = f_1(z_1)\nu_0(K)\boldsymbol{\Delta}_\varepsilon$. By the fact that

$$\boldsymbol{\Omega}^{(1)}(z_1)^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_1 & -\mathbf{M}_1(z_1)'\mathbf{C}\mathbf{J}_1 \\ -\mathbf{J}_1\mathbf{C}'\mathbf{M}_1(z_1) & \mathbf{J}_1 \end{pmatrix},$$

we have

$$\begin{aligned} & f_1^{-2}(z_1)\boldsymbol{\Omega}^{(1)}(z_1)^{-1}\mathbf{D}_1^{*'}\boldsymbol{\Lambda}_4\mathbf{D}_1^*\boldsymbol{\Omega}^{(1)}(z_1)^{-1} \\ & = \begin{pmatrix} \sigma_{g_0}^2(z_1) & \boldsymbol{\Sigma}_{g_0,g}(z_1) \\ \boldsymbol{\Sigma}_{g_0,g}(z_1) & \boldsymbol{\Sigma}_g(z_1) \end{pmatrix} \end{aligned}$$

for some $\boldsymbol{\Sigma}_{g_0,g}(z_1)$. Thus, Corollary 3 holds from Corollary 2. Hence, Corollary 3 is proved. \square

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 research related to this work includes 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: PROOFS OF LEMMAS

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 $\mathbf{C}_\beta = \text{diag}\{1, n^\beta \mathbf{C}\}$ and $\tilde{\mathbf{C}}_\beta = \text{diag}\{1, n^\beta \tilde{\mathbf{C}}\}$.

Lemma A.1. *Let $\boldsymbol{\eta}'_i = \mathbf{z}_{i1}^* \otimes \mathbf{z}_i^*$. Then, under Assumptions A1–A4 and A6, we have*

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n (h^{1/2}\boldsymbol{\eta}'_i u_i K_h(z_{i1} - z_1), h^{1/2}\boldsymbol{\eta}'_i \mathbf{v}'_i \mathbf{g}(z_{i1}) \\ & \quad \times K_h(z_{i1} - z_1), \mathbf{z}_i \mathbf{v}'_i) \Rightarrow (\mathbf{Z}_{ku}, \mathbf{Z}_{kvg}, \mathbf{Z}_v), \end{aligned}$$

where the joint distribution of \mathbf{Z}_{ku} , \mathbf{Z}_{kvg} , and $\text{Vec}(\mathbf{Z}_v)$ is $N(\mathbf{0}, \boldsymbol{\Lambda})$.

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

$$\boldsymbol{\xi}_i = \begin{pmatrix} h^{1/2} \boldsymbol{\eta}'_i u_i K_h(z_{i1} - z_1) \\ h^{1/2} \boldsymbol{\eta}'_i \mathbf{v}'_i \mathbf{g}(z_{i1}) K_h(z_{i1} - z_1) \\ \text{Vec}(\mathbf{z}_i \mathbf{v}'_i) \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{\xi}_{i1} \\ \boldsymbol{\xi}_{i2} \\ \boldsymbol{\xi}_{i3} \end{pmatrix}.$$

Clearly, $E[\boldsymbol{\xi}_i] = \mathbf{0}$. Since $\{(\mathbf{z}_i, u_i, \mathbf{v}_i)\}$ are independent and identically distributed, then, so are $\{\boldsymbol{\xi}_i\}$. It follows from the central limit theorem and the kernel smoothing technique (e.g., Fan and Gijbels (1996)) as well as the Cramér-Wold device that $n^{-1/2} \sum_{i=1}^n \boldsymbol{\xi}_i \Rightarrow N(\mathbf{0}, \boldsymbol{\Lambda})$ holds. The remaining is to show that $\boldsymbol{\Lambda}$ is the limiting covariance matrix of $\boldsymbol{\xi}_i$. Indeed, $\text{Cov}(\boldsymbol{\xi}_i)$ is written as

$$\text{Cov}(\boldsymbol{\xi}_i) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{pmatrix},$$

where $\mathbf{A}_{11} = \text{Var}(\boldsymbol{\xi}_{i1})$, $\mathbf{A}_{12} = \text{Cov}(\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i2})$, $\mathbf{A}_{13} = \text{Cov}(\boldsymbol{\xi}_{i1}, \boldsymbol{\xi}_{i3})$, $\mathbf{A}_{22} = \text{Var}(\boldsymbol{\xi}_{i2})$, $\mathbf{A}_{23} = \text{Cov}(\boldsymbol{\xi}_{i2}, \boldsymbol{\xi}_{i3})$, $\mathbf{A}_{33} = \text{Var}(\boldsymbol{\xi}_{i3})$, $\mathbf{A}_{21} = \mathbf{A}'_{12}$, $\mathbf{A}_{31} = \mathbf{A}'_{13}$, and $\mathbf{A}_{32} = \mathbf{A}'_{23}$. By Assumptions A1–A4 and A6, one can show easily that $\boldsymbol{\eta}'_i \boldsymbol{\eta}_i = \mathbf{z}_{i1}^{*\otimes 2} \otimes \mathbf{z}_i^{*\otimes 2}$, $\mathbf{A}_{33} = E[\boldsymbol{\Sigma}_{vv}(\mathbf{z}_i) \otimes \mathbf{z}_i \mathbf{z}'_i] = \boldsymbol{\Lambda}_3$,

$$\begin{aligned} \mathbf{A}_{11} &= h E \left[\mathbf{z}_{i1}^{*\otimes 2} \otimes \mathbf{z}_i^{*\otimes 2} u_i^2 K_h^2(z_{i1} - z_1) \right] \\ &= f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \boldsymbol{\Delta}_{uu} + o(1), \\ \mathbf{A}_{12} &= h E \left[\mathbf{g}'(z_{i1}) \mathbf{v}_i u_i \mathbf{z}_{i1}^{*\otimes 2} \otimes \mathbf{z}_i^{*\otimes 2} K_h^2(z_{i1} - z_1) \right] \\ &= f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \boldsymbol{\Delta}_{uv} + o(1), \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}_{22} &= h E \left[\mathbf{g}(z_{i1})' \mathbf{v}_i \mathbf{v}'_i \mathbf{g}(z_{i1}) \mathbf{z}_{i1}^{*\otimes 2} \otimes \mathbf{z}_i^{*\otimes 2} K_h^2(z_{i1} - z_1) \right] \\ &= f_1(z_1) \text{diag}\{\nu_0(K), \nu_2(K)\} \otimes \boldsymbol{\Delta}_{vv} + o(1). \end{aligned}$$

Similarly, one can show easily that

$$\begin{aligned} \mathbf{A}_{13} &= h^{1/2} E \left[\begin{pmatrix} u_i v_{i1} & \dots & u_i v_{ip} \\ (z_{i1} - z_1) u_i v_{i1}/h & \dots & (z_{i1} - z_1) u_i v_{ip}/h \end{pmatrix} \right. \\ &\quad \left. \otimes \begin{pmatrix} \mathbf{z}'_i \\ \mathbf{z}_i \mathbf{z}'_i \end{pmatrix} K_h(z_{i1} - z_1) \right] \\ &= h^{1/2} f_1(z_1) E \left[\begin{pmatrix} \boldsymbol{\Sigma}_{uv}(\mathbf{z}_i) \\ \mathbf{0} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{z}'_i \\ \mathbf{z}_i \mathbf{z}'_i \end{pmatrix} \middle| z_{i1} = z_1 \right] \\ &\quad + o(h^{1/2}) = o(1), \end{aligned}$$

and

$$\begin{aligned} \mathbf{A}_{23} &= h^{1/2} E \left[\begin{pmatrix} \mathbf{g}'(z_{i1}) \mathbf{v}_i \mathbf{v}'_i \\ \mathbf{g}'(z_{i1}) \mathbf{v}_i \mathbf{v}'_i \frac{z_{i1} - z_1}{h} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{z}'_i \\ \mathbf{z}_i \mathbf{z}'_i \end{pmatrix} K_h(z_{i1} - z_1) \right] \\ &= h^{1/2} f_1(z_1) E \left[\begin{pmatrix} \mathbf{g}'(z_1) \boldsymbol{\Sigma}_{vv}(\mathbf{z}_i) \\ \mathbf{0} \end{pmatrix} \otimes \begin{pmatrix} \mathbf{z}'_i \\ \mathbf{z}_i \mathbf{z}'_i \end{pmatrix} \middle| z_{i1} = z_1 \right] \\ &\quad + o(h^{1/2}) = o(1). \end{aligned}$$

Therefore, we prove the lemma. \square

Lemma A.2. Under Assumptions A1–A4 and A6,

$$n^{1/2-\alpha} \left[\widehat{\mathbf{C}} - \mathbf{C} \right] \Rightarrow \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v.$$

Moreover, for $c(\alpha) \leq 2$, $\widehat{\mathbf{C}} \Rightarrow \mathbf{D}_{c(\alpha)}$, and for $c(\alpha) = 3$, $n^{1/2-\alpha} \widehat{\mathbf{C}} \Rightarrow \mathbf{D}_{c(\alpha)}$.

Proof. Since $\{\mathbf{z}_i\}$ are iid, it follows by a law of large numbers that $n^{-1} \mathbf{z}' \mathbf{z} \rightarrow^p \boldsymbol{\Sigma}_{zz}$. It follows from equations (4) and (5) that $\widehat{\mathbf{C}} = \mathbf{C} + n^{\alpha-1/2} (n^{-1} \mathbf{z}' \mathbf{z})^{-1} (n^{-1/2} \mathbf{z}' \mathbf{v})$, which, in conjunction with Lemma A.1, implies that $n^{1/2-\alpha} (\widehat{\mathbf{C}} - \mathbf{C}) \Rightarrow \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v$. In particular, for $0 < \alpha < 1/2$, $\widehat{\mathbf{C}} \rightarrow^p \mathbf{C}$. Therefore, Lemma A.2 holds. \square

Lemma A.3. Under Assumptions A1–A7, we have

$$\mathbf{B}_{l,j} \equiv n^{-1} \sum_{i=1}^n \mathbf{z}_i^{\otimes l} \zeta_i^j K_h(z_{i1} - z_1) \rightarrow^p \mu_j(K) f_1(z_1) \mathbf{M}_l(z_1),$$

where $\zeta_i = (z_{i1} - z_1)/h$, $\mathbf{z}_i^{\otimes 2} = \mathbf{z}_i \mathbf{z}'_i$ and $\mathbf{z}_i^{\otimes 1} = \mathbf{z}_i$.

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

Proof of Lemma 1. First, we consider the case when $c(\alpha) \leq 2$. To this end, recall definitions of $\widehat{\boldsymbol{\pi}}_i$ and \mathbf{z}_{i1}^* . Then, it is clear that

$$\widehat{\boldsymbol{\pi}}_i^{\otimes 2} = \mathbf{H}_{c(\alpha)}^{-1} \widehat{\mathbf{C}}_0' \begin{pmatrix} 1 & \mathbf{z}'_i \\ \mathbf{z}_i & \mathbf{z}_i \mathbf{z}'_i \end{pmatrix} \widehat{\mathbf{C}}_0 \mathbf{H}_{c(\alpha)}^{-1} \equiv \mathbf{H}_{c(\alpha)}^{-1} \widehat{\mathbf{C}}_0' \mathbf{L}_i \widehat{\mathbf{C}}_0 \mathbf{H}_{c(\alpha)}^{-1},$$

where the definition of \mathbf{L}_i is apparent, and

$$\begin{aligned} \mathbf{z}_{i1}^{*\otimes 2} \otimes \widehat{\boldsymbol{\pi}}_i^{\otimes 2} &= \begin{pmatrix} \mathbf{H}_{c(\alpha)}^{-1} \widehat{\mathbf{C}}_0' & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{c(\alpha)}^{-1} \widehat{\mathbf{C}}_0' \end{pmatrix} \begin{pmatrix} \mathbf{L}_i & \mathbf{L}_i \zeta_i \\ \mathbf{L}_i \zeta_i & \mathbf{L}_i \zeta_i^2 \end{pmatrix} \\ &\quad \times \begin{pmatrix} \widehat{\mathbf{C}}_0 \mathbf{H}_{c(\alpha)}^{-1} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{C}}_0 \mathbf{H}_{c(\alpha)}^{-1} \end{pmatrix} \\ &= (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}^{-1}) (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0') \begin{pmatrix} \mathbf{L}_i & \mathbf{L}_i \zeta_i \\ \mathbf{L}_i \zeta_i & \mathbf{L}_i \zeta_i^2 \end{pmatrix} \\ &\quad \times (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0) (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}^{-1}). \end{aligned}$$

Therefore, by (7), we can re-write $\widehat{\mathbf{S}}_n$ as follows

$$\begin{aligned} \widehat{\mathbf{S}}_n &= (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}^{-1}) (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0') \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} \\ &\quad \times (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0) (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}^{-1}), \end{aligned}$$

where $\mathbf{B}_0 = \begin{pmatrix} \mathbf{B}_{0,0} & \mathbf{B}'_{1,0} \\ \mathbf{B}_{1,0} & \mathbf{B}_{2,0} \end{pmatrix}$, $\mathbf{B}_1 = \begin{pmatrix} \mathbf{B}_{0,1} & \mathbf{B}'_{1,1} \\ \mathbf{B}_{1,1} & \mathbf{B}_{2,1} \end{pmatrix}$, and $\mathbf{B}_2 = \begin{pmatrix} \mathbf{B}_{0,2} & \mathbf{B}'_{1,2} \\ \mathbf{B}_{1,2} & \mathbf{B}_{2,2} \end{pmatrix}$. Then, it follows from Lemmas A.2 and A.3

and the fact that $\mathbf{M}(z_1) = \begin{pmatrix} 1 & \mathbf{M}_1(z_1)' \\ \mathbf{M}_1(z_1) & \mathbf{M}_2(z_1) \end{pmatrix}$ that for $j \geq 0$,

$$\mathbf{B}_j \Rightarrow f_1(z_1)\mu_j(K)\mathbf{M}(z_1),$$

and

$$\begin{aligned} & (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}) \\ &= \begin{pmatrix} \mathbf{I}_2 & \widehat{\mathbf{C}}_0' \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0) \\ &\Rightarrow f_1(z_1) \begin{pmatrix} \mathbf{I}_2 & \mathbf{D}_{c(\alpha)}^{*'} \\ \mathbf{0} & \mu_2(K)\mathbf{M}(z_1) \end{pmatrix} \begin{pmatrix} \mathbf{M}(z_1) & \mathbf{0} \\ \mathbf{0} & \mu_2(K)\mathbf{M}(z_1) \end{pmatrix} \\ &\quad \times \begin{pmatrix} \mathbf{I}_2 & \mathbf{D}_{c(\alpha)}^* \end{pmatrix} = f_1(z_1)\mathbf{S}^{(c(\alpha))}(z_1). \end{aligned}$$

Now, we consider the case when $c(\alpha) = 3$. Similarly, we have

$$\begin{aligned} & (\mathbf{I}_2 \otimes \mathbf{H}_3)\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_3) \\ &= \begin{pmatrix} \mathbf{I}_2 & \widehat{\mathbf{C}}_{1/2-\alpha}' \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} \begin{pmatrix} \mathbf{B}_0 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix} (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_{1/2-\alpha}). \end{aligned}$$

Then, it follows from Lemmas A.2 and A.3 that

$$(\mathbf{I}_2 \otimes \mathbf{H}_3)\widehat{\mathbf{S}}_n(\mathbf{I}_2 \otimes \mathbf{H}_3) \Rightarrow f_1(z_1)\mathbf{S}^{(3)}(z_1).$$

The proof of Lemma 1 is complete. \square

Proof of Lemma 2. It is easy to see that

$$\mathbf{P}_n = n^{-1} \sum_{i=1}^n \mathbf{z}_{i1}^* \otimes \widehat{\boldsymbol{\pi}}_i K_h(z_{i1} - z_1)(u_i + \mathbf{v}'_i \mathbf{g}(z_{i1})).$$

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

$$\begin{aligned} \sqrt{nh}\mathbf{P}_n &= (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)}^{-1}) (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0') n^{-1/2} h^{1/2} \\ &\quad \times \sum_{i=1}^n \boldsymbol{\eta}'_i K_h(z_{i1} - z_1)(u_i + \mathbf{v}'_i \mathbf{g}(z_{i1})), \end{aligned}$$

where $\boldsymbol{\eta}_i$ is defined in Lemma A.1. Then, from Lemmas A.1 and A.2, we obtain

$$\sqrt{nh}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{P}_n \Rightarrow (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^{*'})(\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}).$$

Next, for the case when $c(\alpha) = 3$, we re-express $\sqrt{nh}\mathbf{P}_n$ as

$$\begin{aligned} \sqrt{nh}\mathbf{P}_n &= (\mathbf{I}_2 \otimes \mathbf{H}_3^{-1}) (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_{1/2-\alpha}') n^{-1/2} h^{1/2} \\ &\quad \times \sum_{i=1}^n \boldsymbol{\eta}'_i K_h(z_{i1} - z_1)(u_i + \mathbf{v}'_i \mathbf{g}(z_{i1})). \end{aligned}$$

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

$$\sqrt{nh}(\mathbf{I}_2 \otimes \mathbf{H}_3)\mathbf{P}_n \Rightarrow (\mathbf{I}_2 \otimes \mathbf{D}_3^{*'})(\mathbf{Z}_{ku} + \mathbf{Z}_{kvg}).$$

This completes the proof of Lemma 2. \square

Proof of Lemma 3. For z_{i1} in a neighborhood of z_1 , by the Taylor expansion,

$$\begin{aligned} g_j(z_{i1}) &= g_j(z_1) + (z_{i1} - z_1)g_j^{(1)}(z_1) \\ &\quad + \frac{1}{2}(z_{i1} - z_1)^2 g_j^{(2)}(z_1) + o_p(h^2). \end{aligned}$$

Then,

$$\begin{aligned} \mathbf{Q}_n &= n^{-1} \sum_{i=1}^n \mathbf{z}_{i1}^* \otimes \widehat{\boldsymbol{\pi}}_i K_h(z_{i1} - z_1) \\ &\quad \times \boldsymbol{\pi}'_i \left[\frac{1}{2}(z_{i1} - z_1)^2 \mathbf{g}^{*(2)}(z_1) + o_p(h^2) \right]. \end{aligned}$$

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

$$\begin{aligned} (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{Q}_n &= \frac{1}{2}h^2 (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0') \begin{pmatrix} \mathbf{B}_4 \\ \mathbf{B}_5 \end{pmatrix} \\ &\quad \otimes \mathbf{D}_1^* \mathbf{H}_{c(\alpha)}^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2), \end{aligned}$$

where $\mathbf{B}_4 = \begin{pmatrix} \mathbf{B}_{0,2} & \mathbf{B}'_{1,2} \\ \mathbf{B}_{1,2} & \mathbf{B}'_{2,2} \end{pmatrix}$ and $\mathbf{B}_5 = \begin{pmatrix} \mathbf{B}_{0,3} & \mathbf{B}'_{1,3} \\ \mathbf{B}_{1,3} & \mathbf{B}'_{2,3} \end{pmatrix}$ with $\mathbf{B}_{l,j}$ defined in Lemma A.3. An application of Lemmas A.2 and A.3 leads to

$$\begin{aligned} \square \quad (\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{Q}_n &= \frac{1}{2}h^2 (\mathbf{I}_2 \otimes \mathbf{D}_{c(\alpha)}^{*'}) \\ &\quad \times \begin{pmatrix} f_1(z_1)\mu_2(K)\mathbf{M}(z_1)\mathbf{D}_1^* \\ \mathbf{0} \end{pmatrix} \\ &\quad \times \mathbf{H}_{c(\alpha)}^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2) \\ &= \frac{1}{2}h^2 f_1(z_1) \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \otimes (\mathbf{D}_{c(\alpha)}^{*'}\mathbf{M}(z_1)\mathbf{D}_1^*) \\ &\quad \times \mathbf{H}_{c(\alpha)}^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2). \end{aligned}$$

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

$$\begin{aligned} (\mathbf{I}_2 \otimes \mathbf{H}_3)\mathbf{Q}_n &= \frac{1}{2}h^2 (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_{1/2-\alpha}') \begin{pmatrix} \mathbf{B}_4 \\ \mathbf{B}_5 \end{pmatrix} \\ &\quad \otimes \mathbf{D}_1^* \mathbf{H}_1^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2) \\ &= \frac{1}{2}h^2 (\mathbf{I}_2 \otimes \mathbf{D}_3^{*'}) \begin{pmatrix} f_1(z_1)\mu_2(K)\mathbf{M}(z_1)\mathbf{D}_1^* \\ \mathbf{0} \end{pmatrix} \\ &\quad \times \mathbf{H}_1^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2) \\ &= \frac{1}{2}h^2 f_1(z_1) \begin{pmatrix} \mu_2(K) \\ \mathbf{0} \end{pmatrix} \otimes (\mathbf{D}_3^{*'}\mathbf{M}(z_1)\mathbf{D}_1^*) \\ &\quad \times \mathbf{H}_1^{-1} \mathbf{g}^{*(2)}(z_1) + o_p(h^2). \end{aligned}$$

This proves Lemma 3. \square

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

$$\begin{aligned} & n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{R}_{n,1} \\ &= (\mathbf{I}_2 \otimes \widehat{\mathbf{C}}_0') \begin{pmatrix} \mathbf{B}_6 \\ \mathbf{B}_7 \end{pmatrix} n^{1/2-\alpha} [\mathbf{C} - \widehat{\mathbf{C}}] \mathbf{g}(z_1), \end{aligned}$$

where $\mathbf{B}_6 = \begin{pmatrix} \mathbf{B}'_{1,0} \\ \mathbf{B}_{2,0} \end{pmatrix}$ and $\mathbf{B}_7 = \begin{pmatrix} \mathbf{B}'_{1,1} \\ \mathbf{B}_{2,1} \end{pmatrix}$. Applying Lemmas A.2 and A.3, we have

$$\begin{aligned} & n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{R}_{n,1} \\ & \Rightarrow -f_1(z_1) \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \otimes \left[\mathbf{D}_{c(\alpha)}^* \begin{pmatrix} \mathbf{M}_1(z_1)' \\ \mathbf{M}_2(z_1) \end{pmatrix} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v \mathbf{g}'(z_1) \right]. \end{aligned}$$

By an analogue,

$$\begin{aligned} & h^{-1}n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_{c(\alpha)})\mathbf{R}_{n,2} \\ & = \left(\mathbf{I}_2 \otimes \widehat{\mathbf{C}}'_0 \right) \begin{pmatrix} \mathbf{B}_7 \\ \mathbf{B}_8 \end{pmatrix} n^{1/2-\alpha}(\mathbf{C} - \widehat{\mathbf{C}})\mathbf{g}'(z_1) \\ & \Rightarrow -f_1(z_1) \begin{pmatrix} \mathbf{0} \\ \mu_2(K) \end{pmatrix} \otimes \left[\mathbf{D}_{c(\alpha)}^* \begin{pmatrix} \mathbf{M}_1(z_1)' \\ \mathbf{M}_2(z_1) \end{pmatrix} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v \mathbf{g}'(z_1) \right], \end{aligned}$$

where $\mathbf{B}_8 = \begin{pmatrix} \mathbf{B}'_{1,2} \\ \mathbf{B}_{2,2} \end{pmatrix}$. Next, for the case when $c(\alpha) = 3$, by the same token,

$$\begin{aligned} & n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_3(n))\mathbf{R}_{n,1} \\ & = \left(\mathbf{I}_2 \otimes \widehat{\mathbf{C}}'_{1/2-\alpha} \right) \begin{pmatrix} \mathbf{B}_6 \\ \mathbf{B}_7 \end{pmatrix} n^{1/2-\alpha}(\mathbf{C} - \widehat{\mathbf{C}})\mathbf{g}(z_1) \\ & \Rightarrow -f_1(z_1) \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \otimes \left[\mathbf{D}_3^* \begin{pmatrix} \mathbf{M}_1(z_1)' \\ \mathbf{M}_2(z_1) \end{pmatrix} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v \mathbf{g}(z_1) \right] \end{aligned}$$

and

$$\begin{aligned} & h^{-1}n^{1/2}(\mathbf{I}_2 \otimes \mathbf{H}_3(n))\mathbf{R}_{n,2} \\ & = \left(\mathbf{I}_2 \otimes \widehat{\mathbf{C}}'_{1/2-\alpha} \right) \begin{pmatrix} \mathbf{B}_7 \\ \mathbf{B}_8 \end{pmatrix} n^{1/2-\alpha}(\mathbf{C} - \widehat{\mathbf{C}})\mathbf{g}'(z_1) \\ & \Rightarrow -f_1(z_1) \begin{pmatrix} \mathbf{0} \\ \mu_2(K) \end{pmatrix} \otimes \left[\mathbf{D}_3^* \begin{pmatrix} \mathbf{M}_1(z_1)' \\ \mathbf{M}_2(z_1) \end{pmatrix} \boldsymbol{\Sigma}_{zz}^{-1} \mathbf{Z}_v \mathbf{g}'(z_1) \right]. \end{aligned}$$

This accomplishes the proof of the lemma. \square

Received 25 October 2007

REFERENCES

- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106** 979–1014.
- Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is Weak. *Journal of the American Statistical Association* **90** 443–450.
- Cai, Z. (2002). Two-step likelihood estimation procedure for varying-coefficient models. *Journal of Multivariate Analysis* **81** 189–209. [MR1918620](#)
- Cai, Z., Das, M., Xiong, H. and Wu, X. (2006). Functional-coefficient instrumental variables models. *Journal of Econometrics* **133** 207–241. [MR2250179](#)
- Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* **95** 941–956. [MR1804449](#)
- Cai, Z., Fang, Y. and Li, H. (2007). Panel estimation of weak instrumental variables models. *Working Paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association* **103**.
- Campbell, J. Y. (2003). Consumption-based asset pricing. *Handbook of the Economics of Finance*, Elsevier, Amsterdam, 801–855.
- Caner, M. and Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. *Econometric Theory* **20** 813–843. [MR2089143](#)
- Chao, J. and Swanson, N. (2007). Alternative approximations of the bias and MSE of the IV estimator under weak identification with application to bias correction. *Journal of Econometrics* **137** 515–555. [MR2354955](#)
- Das, M. (2005). Instrumental variables estimators for nonparametric models with discrete endogenous regressors. *Journal of Econometrics* **124** 335–361. [MR2125370](#)
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London. [MR1383587](#)
- Hahn, J., Hausman, J. and Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher order bias and MSE approximations. *Econometrics Journal* **7** 272–306. [MR2076636](#)
- Hahn, J. and Kuersteiner, G. (2002). Discontinuities of weak instrument limiting distributions. *Economics Letters* **75** 325–331. [MR1950218](#)
- Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B* **55** 757–796. [MR1229881](#)
- Hausman, J., Stock, J. and Yogo, M. (2005). Asymptotic properties of the Hahn-Hausman test for weak instruments. *Economics Letters* **89** 333–342. [MR2180964](#)
- Li, H. (2006). Semiparametric weak instrumental variables models for panel and cross-sectional data. *Ph.D. Dissertation*, Department of Mathematics and Statistics, University of North Carolina at Charlotte.
- Maddala, G. S. and Jeong, J. (1992). On the exact small sample distribution of instrumental variable estimator. *Econometrica* **60** 181–183.
- Mavroeidis, S. (2004). Weak identification of forward-looking models in monetary models. *Oxford Bulletin of Economics and Statistics* **66**, Supplement, 609–635.
- Nason, J. M. and Smith, G. W. (2005). Identifying the new Keynesian Phillips curve. *Working Paper 2005-1*, Federal Reserve Bank of Atlanta.
- Neeley, C. J., Roy, A. and Whiteman, C. H. (2001). Risk aversion versus inter-temporal substitution: A case study of identification failure in the inter-temporal consumption capital asset pricing Mmodel. *Journal of Business and Economic Statistics* **19** 395–403. [MR1963374](#)
- Nelson, C. and Startz, R. (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* **58** 967–976. [MR1064853](#)
- Newey, W. K., Powell, J. L. and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica* **67** 565–603. [MR1685723](#)
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selection for local least squares regression. *Journal of the American Statistical Association* **90** 1257–1270. [MR1379468](#)
- Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica* **65** 556–586. [MR1445622](#)
- Stock, J. (2002). Instrumental variables in economics and statistics. *International Encyclopedia of the Social Sciences*, Elsevier, Amsterdam, 7577–7582.
- Stock, J. H. and Wright, J. H. (2000). GMM with weak instruments. *Econometrica* **68** 1055–1096. [MR1779144](#)
- Stock, J. H., Wright, J. H. and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **4** 518–529. [MR1973801](#)
- Woodford, M. D. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ.

Yogo, M. (2004). Estimating the elasticity of inter-temporal substitution when instruments are weak. *Review of Economic Studies* **86** 797–810.

Zongwu Cai
Department of Mathematics & Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223, USA

Wang Yanan Institute for Studies in Economics
Xiamen University, China
E-mail address: zca@uncc.edu

Henong Li
Department of Mathematics & Statistics
University of North Carolina at Charlotte
Charlotte, NC 28223, USA